LIBRARY
Education
MAY 9 1960

Vol. 51 APRIL 1960 No. 2

The Journal of Educational Psychology

CONTENTS

Verbal Comprehension and Communication in Negro and White Children..... 47
ARNOLD S. CARSON AND A. I. RABIN

The Social Desirability Variable and Verbal Learning Performance..... 52
PHILIP K. BROWN

Relationships among Several Measures of Pupil Adjustment..... 60
IRA J. SEMLER

The Effect of Teacher-Pupil Contacts Involving Praise on the Sociometric Choices of Students..... 65
NED A. FLANDERS AND SULO HAVUMAKI

The Joplin Plan of Reading VS. A Traditional Method..... 69
ELMER F. MORGAN, JR. AND GERALD R. STUCKER

A Note on the Teaching of Skinner's Descriptive Behaviorism in the Fifth Grade. 74
H. E. KLUGH, W. A. DETERLINE, AND D. E. HENDERSON

Attitude Changes of Student Teachers and the Validity of the Minnesota Teacher Attitude Inventory..... 76
OWEN SCOTT AND STERLING G. BRINKLEY

A Study of Mediation in Mentally Retarded and Normal School Children..... 82
GERSHON BERKSON AND GORDON N. CANTOR

Comparability of Peabody Picture Vocabulary Test Scores under Group and Individual Administration..... 87
RAYMOND C. NORRIS, JOHN V. HOTTEL, AND SADYE BROOKS

Sources of Tension in Married and Single Women Teachers of Different Ages.. 92
MARVIN POWELL AND CHARLES D. FERRARO

AMERICAN PSYCHOLOGICAL ASSOCIATION

The Journal of Educational Psychology

RAYMOND G. KUHLEN, Editor
Syracuse University

Advisory Editors

LEE J. CRONBACH
University of Illinois
FRANCIS J. DIVESTA
Syracuse University
CHESTER W. HARRIS
University of Wisconsin
ERNEST R. HILGARD
Stanford University
NICHOLAS HOBBS
George Peabody College For Teachers
HAROLD E. JONES
University of California, Berkeley
IRVING LORGE
Teachers College, Columbia University
DOROTHEA MCCARTHY
Fordham University
T. ERNEST NEWLAND
University of Illinois
H. H. REMMERS
Purdue University
JULIAN C. STANLEY
University of Wisconsin
J. B. STROUD
State University of Iowa
DONALD E. SUPER
Teachers College, Columbia University
PERCIVAL M. SYMONDS
Teachers College, Columbia University
GEORGE G. THOMPSON
Ohio State University

The Journal of Educational Psychology publishes original investigations of problems of learning, teaching, and the psychological development of the individual. It contains articles on the following subjects: the psychology of school subjects; experimental studies of learning; the development of interests, attitudes, personality, social relations, and vocational orientation, particularly (but not exclusively) as related to school adjustment or the impact of the school program upon development and adjustment; emotion, motivation, and character; mental development; and methods. The last includes tests, statistical techniques, and research techniques in cross-sectional and developmental studies. Studies published deal with all levels of education—preschool, elementary, secondary, college, and adult education—and with all age groups.

Manuscripts and correspondence on editorial matters should be addressed to the Editor: Raymond G. Kuhlen, Department of Psychology, Syracuse University, 123 College Place, Syracuse 10, New York.

All manuscripts should be submitted *in duplicate* to facilitate referral to advisory editors and must be prepared to conform to the style and procedures described in the *Publication Manual of the American Psychological Association*. Because of the great demand for journal space, authors should give special attention to reducing the length of manuscripts while maintaining clarity of expression. Ordinarily, papers should not exceed 15 typed pages. Accepted manuscripts are published in order of their receipt. An author may, however, request early publication of accepted manuscripts if he is willing to assume the cost of printing the necessary extra pages.

ARTHUR C. HOFFMAN	HELEN ORR	FRANCES NOYES
Managing Editor	Promotion Manager	Editorial Assistant

Communications including subscriptions, orders of back issues, and changes of address should be addressed to the American Psychological Association, 1333 Sixteenth Street, N.W., Washington 6, D. C. Address changes must reach the APA by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Annual subscription: $8.00 (Foreign $8.50). Single copies, $1.50.

PUBLISHED BIMONTHLY BY:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Mt. Royal & Guilford Aves., Baltimore 2, Md.
and 1333 Sixteenth Street N.W., Washington 6, D.C.

Second-Class postage paid at Baltimore, Md.

Copyright, 1960 by the American Psychological Association, Inc.

THE JOURNAL OF
EDUCATIONAL PSYCHOLOGY

Volume 51 April 1960 Number 2

VERBAL COMPREHENSION AND COMMUNICATION IN NEGRO AND WHITE CHILDREN[1]

ARNOLD S. CARSON AND A. I. RABIN[2]
Michigan State University

Comparisons of intellectual functioning in Negroes and whites have been of concern to research workers for some time. The classical review and summary by Klineberg (1935) is not yet outdated. For the most part, the superiority of the whites, especially on verbal tasks, has been demonstrated. However, the improvement in intelligence test scores of Negro children who migrated to the North from Southern states has also been shown. Moreover, the degree of improvement tends to be related to duration of residence in the North.

Performance tasks which correlate highly with "general intelligence", such as the Goodenough Draw-A-Man IQ Test, have not yielded significant differences between Negro and white children (Anastasi & D'Angelo, 1952). It would appear that tasks requiring verbal comprehension and verbal expression are the ones that are to a large extent responsible for the significant differences between the groups.

The present study is concerned with the investigation of these two functions, comprehension and expression or communication, in Negro and white children. We propose to study these functions separately, and in relation to each other, in Southern and Northern Negro, and in Northern white, children. Our prediction was, in consonance with previous related findings, that when groups are matched for age, grade placement, sex, and level of verbal comprehension, the white children will be superior to the Southern and Northern Negroes in verbal communication; also, that the Northern Negro children will be superior to the Southern Negro children on measures of verbal communication.

METHOD

The Full Range Picture Vocabulary Test (Ammons & Ammons, 1949) was selected as the method of determining the level of verbal comprehension. This test requires no verbal response from the subjects (*S*). It consists of a series of cards with several pictures on each card. The *S* is merely required to point to the specific picture each time the examiner reads a word from the vocabulary test.

Two measures of verbal communication were employed for comparative purposes. First, the WISC Vocabulary (Weschsler, 1949) was presented in the standard manner and the responses scored as provided in the manual. Second, the words of the Full-Range Picture Vocabulary Test (FRPVT) which were "defined" by gesticulation were presented as a conventional vocabulary test, requiring oral definition.

The conventional scoring of vocabulary tests is an "all or none" affair. No provisions are made for qualitative differences in definitions or in levels of abstraction,

[1] Based on master's thesis (Carson, 1959) submitted to the Department of Psychology, Michigan State University by A. S. Carson and supervised by A. I. Rabin.

[2] The authors are grateful to G. F. King and C. F. Wrighley for their valuable assistance and suggestions.

precision and communication. This lack of sensitivity in the conventional scoring of vocabulary tests has been pointed out long ago by Yacorzynski (1940).

In order to provide for scoring of levels of communication with the vocabulary of the FRPVT an adaptation of a method reported in the literature (Rabin, King, & Ehrmann, 1955) was utilized. A description of this Qualitative Vocabulary Scale, of the categories of responses and definitions of the word "wagon" as illustrative of the levels of communication appear below.

Class 1. *Categorization and Synonym*

a. Categorization. The categorization responses classified by some definite scheme in terms of its universal characteristics.

Response—"a vehicle"

b. Synonym. The synonym response may essentially be used to replace the object or idea with no or little change in the denotative aspects of the stimulus word.

Response—"a cart"

Class 2. *Essential Description.* An essential description response must give the characterizing features of the stimulus word. If the stimulus word is abstract, the response must create mental imagery of the relevant situation. If the stimulus word is concrete (physically tangible), the response must differentiate between the stimulus word and members in its class.

Response—"It's a wooden thing with 4 wheels, and it looks like a box."

Class 3. *Essential Function.* An essential function response must describe primary rather than peripheral usage or purpose of an object or an idea.

Response—"You ride in it out West."

Class 4. *Example.* An example response defines an object or idea in terms of its aspects or members.

Response—"There's a red wagon kids play with."

Class 5. *Vague Description and Vague Function*

a. Vague Description. A vague description is a response that is not totally irrelevant but does not give the characterizing features of the object or idea.

Response—"Something that has four wheels."

b. Vague Function. A vague function response describes the peripheral rather than primary usage or purpose of the object or the idea.

Response—"It bumps into people."

Class 6. *Error.* The error response is totally irrelevant to the stimulus word.

Response—"The dog 'wagons' his tail."

Class 7. *Don't Know.* A "don't know" response is a statement or a lack of statement designating that the *S* is unable to verbally define a word *which he had previously designated recognition* for by a gesticulation response.

Response—"Don't know."

Subjects

The *Ss* were Southern Negro (SN), Northern Negro (NN), and Northern white (NW) children in the fourth, fifth, and sixth grades. None of the *Ss* were advanced or behind in their age-grade placement. The SN group comprised all of the Negro children within the age range from 9.5 to 11.5 years who were born and reared in the South and had migrated to Lansing, Michigan within 28 months of the date tested. All of the NW and NN children were born and reared in Ingham County, Michigan. Data was collected on the occupational levels of the NW and NN children in accordance with the *Dictionary of Occupational Titles* (United States Government Printing Office, 1944) classification. Two of the main wage earners in the NW children's families were professional, technical or managerial workers, 4 were clerical or sales workers, 3 were service employees, 9 were mechanical workers, and 12 were manual laborers; respectively, there were 0, 5, 2, 5, and 18 NN main wage earners. It may be noted that the occupational levels of main wage earners from the two samples are similar. Due to the recent relocation of the SN families, the main wage earners' vocations were unsettled and therefore unavailable as data.

Children were tested with the FRPVT at random from the NW and NN groups

TABLE 1

GRADE, SEX, AGE, AND MEAN PICTURE VOCABULARY TEST SCORES OF THE THREE GROUPS

Group	N	Grade			Sex		Ages (Months)		Raw Score of FRPVT	
		4th	5th	6th	Male	Female	Mean	SD	Mean	SD
NW	30	16	8	6	16	14	128	10.9	43.67	4.95
NN	30	16	8	6	15	15	125	10.9	43.57	5.74
SN	30	16	8	6	14	16	128	13.5	43.53	5.21

in order to acquire *Ss* whose scores could be individually matched with the scores of children in the SN group. Table 1 summarizes relevant data for the three groups of children.

PROCEDURE

The FRPVT and the two oral vocabulary devices were administered individually to the *Ss* by the same examiner in single testing sessions. The SN children were tested first followed by random testing of NN and NW children who would qualify for the respective samples. When each SN child was matched with a NW and NN child in terms of the FRPVT scores, the testing was discontinued.

Two independent scorers tallied verbal responses to the FRPVT. These responses were scored according to the classification system described above. The total percentage of interscorer agreement was .79. Complete agreement between the two scorers was obtained after the debatable responses were discussed.

RESULTS AND DISCUSSION

As a preliminary step in the treatment of the data, all *Ss* were classified into "high" and "low" responders. The *Ss* whose majority of responses were tallied in categories 1, 2 and 3 were placed in the former classification; the predominance of the remaining four categories characterized the second group. Table 2 reports the incidence of members of each sample in the "high" and "low" categories. The differences in the distribution are statistically significant.

As predicted, the white children assume the top position. It is also interesting to note that the NN group place in an intermediate position between the white and SN children.

The responses of the three groups were also compared on each of the seven categories comprising our qualitative (levels of communication) scoring system. Table 3 summarizes the levels of significance of the differences between the groups computed by means of the Wilcoxon matched-pairs signed-rank test (Siegel, 1956). Thirteen of the 21 comparisons were significant in the predicted direction. The table indicates that the general prediction is supported to a considerable degree. This is especially true with the extremes of the qualitative categories.

The results of the study demonstrated that NW children, of comparable nonverbal word recognition (or comprehension) abilities, manifest higher levels of verbal communication than NN children, and, in turn, the NN children manifest higher levels of verbal communication than SN children.

All comparisons by means of the WISC vocabulary subtest yield significant differences in the predicted direction (using Wilcoxon's test mentioned above).

Again, global vocabulary scores, as in the case of the more refined treatment of the FRPVT, indicate the superiority of the white children compared with the two Negro samples. Also, the NN group is in-

TABLE 2

A COMPARISON OF THE INCIDENCE OF HIGH AND LOW RESPONDERS ON THE QUALITATIVE VOCABULARY SCALE IN THE THREE GROUPS

Group	High	Low	z^1	p^1
NW	25	5		
NN	18	12		
SN	7	23	22.54	$<.001$

[1] One-tailed test.

TABLE 3

SUMMARY OF LEVELS OF SIGNIFICANCE OBTAINED FROM COMPARISONS OF RANKS ON THE QUALITATIVE VOCABULARY SCALE

Class	Groups											
	NW vs. SN (T) (T')				NW vs. NN (T) (T')				NN vs. SN (T) (T')			
	T	T'	z	p	T	T'	z	p	T	T'	z	p
1	271.0	29.0	3.44	<.0005	289.0	62.0	2.87	<.0025	181.0	72.0	1.76	<.05
2	327.5	107.5	2.36	<.01	283.5	122.5	1.82	<.05	206.5	171.5	.41	ns
3	350.0	85.0	2.85	<.0005	233.0	118.0	1.45	ns	326.0	139.0	1.91	<.05
4	274.5	160.5	1.22	ns	126.5	198.5	.82	ns	301.0	105.0	2.22	<.025
5	154.5	223.5	.82	ns	147.0	231.0	1.00	ns	202.5	203.5	.00	ns
6	34.5	316.5	3.57	<.0005	57.5	218.5	2.43	<.01	55.5	295.5	3.04	<.002
7	3.0	432.0	4.63	<.0001	95.5	229.5	1.79	<.05	40.5	394.5	3.82	<.000

[1] If a two-tailed test were employed, the difference would have been significant in a direction contrary to the current hypothesis.

TABLE 4

SUMMARY OF COMPARISONS OF RANKS ON THE WISC VOCABULARY SUBTEST

Comparisons	Total Ranks			Theoretical			Significance
	NW	NN	SN	Mean	SD	z	
NW vs. SN	432.0	—	3.0	217.5	46.25	4.62	<.0001
NN vs. SN	—	345.5	60.6	203.0	43.92	3.24	<.001
NW vs. NN	334.0	101.0	—	217.5	46.25	2.51	<.01

ferior to the NW but superior to the SN children. This finding is consistent with the results in the previous two tables.

The children's pattern of responding is interesting in that it was unique for each of the groups and may therefore be a clue to differences in characteristic thinking among the three groups. The NW children favor the higher levels of verbal communication (Classes 1, 2, 3). Their percentage of responses decreases sharply in the Example classification and after a slight increase in the Vague classification, their percentage of responses continues to decrease lower than that of the two Negro groups in the Error and Don't Know classifications. NN children favor the Vague type response and manifest a greater percentage of Error type responses than the white children. The SN children's pattern is the only one which manifests a high peak in the Don't Know classification.

The differences among the groups cannot be consistently explained according to one theoretical viewpoint. Differences between the NW and NN children could lend support to the contention that constitutional racial differences in intelligence exist or to the viewpoint that a difference of cultural opportunities accounts for the variance between racial groups. However, there are sufficient possibilities of differences in the cultural milieu of the two samples to question their equivalence of social opportunities. The two groups remain segregated from each other in activities other than those which revolve around school. The NN sample, for the most part, lives in a section of town which is a homogeneous Negro settlement. Their cultural milieu is definitely not characteristic of white middle-class modes of living.

The comparisons of the NN and SN children lend themselves to a more clear-cut

interpretation since the cultural variable of geographical residence was considered in the experimental design. It is contended that the superior educational environment and the greater opportunity for cultural advancement of the NN child over the SN child accounts for the significant differences between the two groups.

The present study supports Klineberg's (1947) research findings that geographical residency of the Negro child is an important determinant of vocabulary performance. Coppinger and Ammons' (1952) contention that different norms should be utilized when making intergroup comparisons involving Negroes where the members of the groups have different cultural backgrounds is also supported.

Finally, it should be pointed out that verbal comprehension and verbal communication seem to be two different functions. Although the groups were equated on comprehension, they showed marked differences with respect to communication. The former task requires recognition within a certain context, which is quite different from verbal communication and definition of a word in isolation, and not in context. It would appear that differences between Negroes and whites on conventional intelligence tests, and especially on vocabulary subtests, may be primarily due to failure in verbal communication rather than in comprehension.

Summary

Three groups (30 in each group) of NW, NN, and SN school children, matched for age, sex, grade placement, and level of verbal comprehension, were compared on two vocabulary tests requiring verbal communication. In accord with the original prediction, the white children were superior to the Negro children, and the NN children were superior to the SN children on these two measures. The results were discussed in relation to the possible racial and cultural geographic factors involved.

REFERENCES

Ammons, R. B., & Ammons, H. S. The Full-Range Picture Vocabulary Test. *Amer. Psychologist*, 1949, **4**, 267–268.

Anastasi, Ann, & D'Angelo, Rita Y. A comparison of Negro and white preschool children in language development and Goodenough Draw-A-Man IQ. *J. genet. Psychol.*, 1952, **81**, 147–165.

Carson, A. S. Verbal comprehension and verbal communication in Negro and white children. Unpublished master's thesis, Mich. State Univer., 1959.

Coppinger, N. W., & Ammons, R. B. The Full-Range Picture Vocabulary Test: VIII. A normative study of Negro children. *J. clin. Psychol.*, 1952, **8**, 136–140.

Klineberg, O. *Race differences*. New York: Harper, 1935.

Klineberg, O. Negro intelligence and urban residence. In T. M. Newcomb & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1947. Pp. 24–32.

Rabin, A. I., King, G. F., & Ehrmann, J. C. Vocabulary performance of short-term and long-term schizophrenics. *J. abnorm. soc. Psychol.*, 1955, **50**, 255–258.

Siegel, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

United States Government Printing Office, Division of Occupational Analysis. *Dictionary of occupational titles*. Part IV. Washington, D.C.: Author, 1944.

Wechsler, D. *Wechsler Intelligence Scale for Children*. New York: Psychological Corp., 1949.

Yacorzynski, G. K. An evaluation of the postulates underlying the Babcock deterioration test. *Psychol. Bull.*, 1940, **37**, 425–426. (Abstract).

(Received August 11, 1959)

Journal of Educational Psychology
Vol. 51, No. 2, 1960

THE SOCIAL DESIRABILITY VARIABLE AND VERBAL LEARNING PERFORMANCE[1]

PHILIP K. BROWN[2]

Sonoma State Hospital, Eldridge, California

This article presents a study of the relation between social desirability as a psychological variable and learning performance and recall of college students on a complex verbal learning task.

As developed by Edwards (1953a), the concept of social desirability has two aspects. It may be considered as a *property of the items* in a personality scale such as the Edwards Personal Preference Schedule (EPPS). For example, it seemed obvious that a statement such as "I like to be loyal to my friends" reflected a more socially desirable attitude, and would be rated higher on a dimension of social desirability, than a statement such as "I like to avoid my responsibilities and obligations." Edwards (1953b) has shown that subjects tend to endorse the statements of higher scale value as characteristic of themselves more often than the statements rated low on the social desirability dimension. It follows that the social desirability variable is influential in affecting test responses and would need to be controlled in research using personality scales.

The second aspect of social desirability appeared as an individual differences variable that reflected the tendency of a person to endorse the more socially desirable items as characteristic of himself. As such it may be considered a *property of the person*. Edwards developed the Social Desirability Scale (the SDS) to measure the tendency of a person to "try to look good" in a social situation. The extent to which this second aspect of social desirability influenced test behavior was reported by Fordyce (1956) who found that the SDS correlated at a significantly high level with *every* clinical scale of the MMPI. The SDS accounted for the major portion of the variance of the clinical scales, leaving a minor portion of the variance to be accounted for by the pathology of the person or by other determinants. If the SDS measures an important determinant of test behavior, would it also show a relationship to such behavior as is found in a learning task? This study addressed itself to this question.

There has been a dearth of studies relating the SDS to learning performance. However, the SDS was found to have a very high negative correlation with the Taylor Manifest Anxiety Scale (the A scale) (Edwards, 1957), so that it was felt that results obtained from learning studies which related the A scale to learning behavior, could be applied, with caution, to the SDS. Taylor (1951) first related the A scale to eyelid conditioning, Lucas (1952) to serial learning, Ramond (1953) to complexity of verbal material, and others showed a significant relationship to exist between the A scale and performance.

Some investigators used personality scales that correlated highly with the A scale, but obtained results that were inconsistent with the findings of the previous investigators (e.g., Lazarus, Deese, & Hamilton, 1954; Heilizer, Axelrod, & Cowen, 1954). Such disagreement constituted part of the justification for making this study.

[1] An extended report of this study may be obtained from the Library of Congress, Card No. Mic. 58-1074, for a fee of $2.00, Univer. Microfilms, 313 N. First St., Ann Arbor, Michigan.

[2] This study was based on a doctoral dissertation completed at the University of Washington in 1957. The author wishes to express his thanks to C. R. Strother and the other members of his advisory committee for their help.

Though Taylor and Spence (1952) considered the A scale to measure "drive level" and used it to predict behavior within a Hullian framework, Edwards preferred to think of the SDS as measuring the tendency to "try to look good" in a social situation. Edwards actual definition of the meaning of social desirability is operationally determined by the score obtained and the correlation of the SDS with other personality scales and measures. However, this informal description has been used by him and serves our purpose well.

Statement of the Problem

The aims of this study were to explore the relationship between learning performance and (*a*) the social desirability rating of statements, (*b*) need for social desirability as a personality variable, and (*c*) the interaction of social desirability, learning performance, and sex.

Procedure

The *Ss* in this experiment were required to memorize 16 statements paired to nonsense syllables. Learning proceeded by the modified paired-associates method in which each item learned to the criterion of exact recall on two successive trials was removed. Each trial consisted of the presentation of all unlearned items. The *Ss*' success or failure was recorded each time an item was presented, and this record represented a detailed account of his learning performance. The *Ss* were also asked to recall the entire set of 16 statements, starting one minute after completion of learning, to provide a score for immediate recall. The *Ss* returned after a 24-hour delay and attempted recall again to provide a delayed recall score. This completed their task.

Selection of Items

The items used in the learning task were taken from the 140 statements found in the EPPS. The range of scale values for the statements varied from 0.60 for the statement rated least socially desirable to a scale value of 4.14 for the highest rated item. Each list learned by an *S* was composed of eight items from the high end and eight items from the low end of the SDS. The low scale value items (Low S-V) had a mean scale value of 1.49 as compared to a mean scale value of 3.44 for the items of high scale value (High S-V). The number of words varied in the statements between 8 and 16, with a mode of eleven words. There was no significant difference in mean number of words between the High S-V and Low S-V items. The 16 items were typed on 3×5 cards.

A list of 16 four-letter nonsense syllables were obtained from Woodworth (1938), modified slightly for use as stimulus syllables and were paired with the statements.

Selection of Subjects

Ss were students from beginning classes in psychology at the University of Washington. Students in these classes were given the 39-item SDS and were asked to volunteer for a research project. Some pressure to cooperate existed, since one of the course requirements was to participate in some ongoing research. Final selection of *Ss* was made from among those who had volunteered and whose SD score fell at the extremes of the class distribution of SD scores. About one-fourth of the 600 students who took the SDS volunteered. No significant difference in mean SD score was found between the volunteering and nonvolunteering students.

Four experimental groups of 10 *Ss* each were formed as follows: High SD males, Low SD males, High SD females, and Low SD females. The range and means of SD scores and *Ss* age is shown in Table 1. The mean age of the males is somewhat higher than that of the females due to the presence of several veterans among the male groups.

Experimental Method

Each *S* was contacted by phone and an appointment made for individual testing.

TABLE 1
RANGE AND MEANS OF SD SCORES AND AGE OF *Ss*

Subjects (N = 10)	SD Range	Mean SD Score	Age Range[a] (in years)	Mean Age
High SD males	35–39	36.3	18–32	21.8
Low SD males	22–29	25.8	18–28	21.0
High SD females	32–39	36.4	18–20	18.5
Low SD females	22–28	24.9	18–21	19.6

[a] The males ranged higher in age due to the presence of several veterans in each male group.

When the *S* arrived at the designated room he was seated at a desk facing the *E* and a brief explanation of the purpose of the study was made as follows:

We all know that intelligence isn't the only thing that's important in learning. Some bright people can't seem to learn, and some not-so-bright people are pretty good in school. We've found that certain psychological variables seem to be important in learning and remembering. We're trying to find out more about these variables.

The *S* was then shown a card containing a list of nonsense syllables while the *E* explained the purpose of using them as stimuli in a paired-associates learning task, and the *S*'s responsibility in the experiment.

On the first two trials, the *E* read the nonsense syllable aloud and showed the statement paired with it to the *S*. The *S* repeated the nonsense syllable aloud and read the statement aloud immediately afterward. The cards were shuffled between trials to eliminate the possibility of serial learning. Ten seconds per card was found adequate during the first two trials to allow the *S* time to read it. A 20-second interval was used between trials to allow the *E* time to shuffle the pack. Each card was presented to the *S* by the *E* placing it on the table in front of him. Some variation in the length of exposure occurred for the different cards since it took longer to read some cards than others.

Before starting the third trial, the *S* was asked to guess how many trials it would take him to complete the learning task. His estimate (henceforth, the self confidence score) was taken as a measure of the level of self confidence with which he undertook the task. Starting with the third trial, the *S* was told:

I will say the nonsense syllable aloud and allow you five seconds to recall the statement. If you don't recall it, I will show it to you, and you may read it aloud, and we will continue as before.

Learning then proceeded by the modified paired-associates learning method until all cards had been eliminated. When the task was completed *E* engaged the *S* in small talk for one minute and then asked him to undertake recall of the list. The *E* gave the nonsense syllable and the *S* attempted recall of the statement paired with it. The statements were recalled in the same order in which they had been learned in order to correct in a small way the effect of temporal contiguity of the last learned items. All statements that were not exactly recalled were written down verbatim so that such items of doubtful precision of recall could be later evaluated by independent judges and accepted or rejected.

Upon completion of immediate recall the *S* was instructed to return the next day at the same hour for the second half of the experiment. No other information was given him.

At the second session, the *S* was given a few moments to become acclimated. He was then informed that the *E* wanted to see how much of yesterday's learning he had retained. The nonsense syllables were given in random order. Again, all statements of doubtful precision of recall were recorded verbatim by the *E*.

Thus the *S*'s success or failure on each item presentation was recorded, and provided a score representing the number of trials necessary to master each item. The number of items recalled at the end of

the learning task was taken as the score for immediate recall, and the number of items recalled after 24 hours was taken as the score for delayed recall.

The items of doubtful precision of recall in both immediate and delayed recall were made up into a list in which each doubtful item was paired with its exact statement. The pairs were reviewed by four graduate psychology students acting as judges. They were required to judge whether a doubtful item was sufficiently like its exact statement so that the *S* could be given credit for it. The criterion of acceptance by the *E* was that three of the four judges agree it should be accepted. There was a total of 278 instances to be judged, and of these only 29.1% met the *E*'s criterion. The average agreement between judges was 80.8%.

Results

Table 2 shows the means and sigmas of the self confidence scores for the four experimental groups and gives the *t* tests of significance between the High SD vs. Low SD males and the High SD vs. the Low SD females. No comparisons are shown between the sexes, or between the High SD *Ss* vs. the Low SD *Ss*, as no significant differences between means was found for these groupings.

Logarithmic transformations to correct for proportionality between means and sigmas shown in Table 2 resulted in improving the significance of the mean difference between the two female groups from the .05 level to the .01 level. The self-confidence scores of the High SD females, as estimated by their guessing how many trials it would take to master the learning task, ranged between 5–20 trials. The Low SD females self-confidence scores ranged between 9–100 trials, with half the Low SD females estimating it would take them 20 or more trials to master the learning task. The two male groups showed no significant difference in mean self-confidence scores. Over the 40 *Ss*, a nonsignificant correlation, $r = +.14$, was found between the *Ss*' self confidence scores and their actual number of trials to mastery of the learning task.

TABLE 2

Means, Sigmas, and *t* Tests For Self-Confidence Scores ($N = 10$ per group)

Group	Mean Trials Estimated	Sigma	*t* Tests
High SD males	19.3	12.7	0.02
Low SD males	19.4	9.8	
High SD females	9.0	5.1	2.19[a]
Low SD females	29.6	29.3	

[a] Since the raw score means and sigmas were proportional for the two female groups, logarithmic transformations were employed. This technique resulted in a $t = 3.15$, which was significant at the .01 level. The tabled value above, based on raw scores was significant at the .05 level.

If a High SD score meant that the *S* had a greater need to "try to look good," this tendency was found only in the females and not the males. In fact, throughout the entire study, the male groups as a whole did not differ from the female groups in the mean performance scores on any variable studied. Likewise, on the whole, the High SD *Ss* failed to differ significantly from the Low SD *Ss* in mean scores on any performance variable studied. This problem will be reviewed further in the Discussion section.

An analysis of variance (Table 3) based on the design for repeated measurements was made of the number of item presentations in order to test the null hypothesis of no difference in the number of item presentations to mastery of the 16-item list. No significant difference in scores was found for the "between groups" variable. This was interpreted to mean that the four groups found the task to be essentially equal in difficulty. However, the "between item types" variable was found to be sig-

TABLE 3
ANALYSIS OF VARIANCE OF NUMBER OF ITEM PRESENTATIONS TO MASTERY OF 16-ITEM LIST

Source	*df*	Mean Square	*F* Test
Between groups (a)	3	724.7	1.47
Between item types (b)	3	411.6	5.90**
Groups × item types (b)	9	26.7	—
Between *Ss*, same group (Error term to test correlated scores, [a] above)	36	493.8	
Interaction: Pooled *Ss* × item types (Error term to test uncorrelated scores, [b] above)	108	69.7	

** Significant at the .01 point.

nificant at the .01 point. Over all *Ss*, the High S-V items were mastered in significantly fewer presentations than the lower S-V items. The explanation was found in an evaluation of which item types, e.g., High S-V, the *Ss* attempted to learn at the beginning of the task, and which item types were left for the end. Since the items were randomized before each trial, the *Ss* could give preference to any item type. A tally of the item types mastered at the beginning of the learning task showed that the *Ss* gave priority to high S-V items. The actual ratio was 1.96:1.00 favoring the High S-V items over the Low S-V items. The ratio for the last learned item types showed the reversed trend. The Low S-V items were mastered in the ratio of 1.76:1.00 over the High S-V items. It seemed clear that the *Ss* selected the High S-V items for learning early in the task.

The actual mean trials to mastery of the 16-item list is shown in Table 4. The mean trials was computed by dividing each *S*'s total number of item presentations by 16, the number of items. It can be seen that for the two female groups, a significant difference in mean trials to mastery was reached favoring the High SD females. The difference in mean trials between the male groups failed even to approach significance. The High SD *Ss* did not differ significantly from the Low SD *Ss* in their mean trials to mastery, nor did the males differ from the females. These values were excluded from the table.

An analysis of variance based on the design for repeated measurements was done on the scores for immediate recall (table not shown) to test the null hypothesis of no difference in scores for immediate recall. Neither of the main variables or their interaction accounted for a significant proportion of the variance. A similar analysis of variance for the scores for delayed recall (table not shown) failed to show any variable accounting for a significant proportion of the variance. Therefore, for immediate and delayed recall we are unable to reject the null hypothesis of no difference in performance scores.

Table 5 shows the mean immediate recall scores for the four experimental groups. It can be seen that when the intrasex mean scores are compared, the males showed no significant difference in immediate recall, while the difference in mean scores between the female groups was significant at the .05 level, in favor of the High SD females.

Table 6 shows the mean delayed recall scores for the four experimental groups

TABLE 4
MEANS, SIGMAS, AND *t* TESTS FOR TRIALS TO MASTERY
(N = 10 per group)

Group	Mean Trials	Sigma	*t* Tests
High SD males	20.3	4.34	
			0.76
Low SD males	18.7	5.01	
High SD females	16.9	4.15	
			2.49*
Low SD females	21.7	4.50	

* Significant at the .05 level.

It can be seen that the High SD females again significantly outclass the Low SD females, while the males showed no significant difference in mean delayed recall scores between their two groups. As before, no significant difference was found between males and females, or between the High SD *Ss* and the Low SD *Ss*.

In summary, the High SD females entered the learning task with a more confident attitude than the Low SD females, learned the 16-item list in significantly fewer trials, showed better immediate recall, and maintained their superiority over a 24-hour delay. The two male groups did not differ significantly from each other on any performance variable. Though not shown in the tables, the males did not differ in mean score from the females on any variable, nor did the High SD *Ss* differ from the Low SD *Ss* on any performance variable.

Discussion

Taylor and Spence (1952) originally considered the A scale as a measure of drive level, and used it to predict behavior of *Ss* in accordance with the Hullian formulation: ${}_sE_R = H \times D$; where ${}_sE_R$ was the reaction potential, H was the habit or learning strength, and D was the measure of drive level. It can be seen that by holding H constant, ${}_sE_R$ became a function of D, the drive level. It could then be predicted that *Ss* with a high *A* score (high drive level) would perform better than *Ss* with a low *A* score. Since the high *A* score is equivalent to a Low SD score, due to the high negative correlation between the two scales, we would expect Low SD *Ss* to do better on our learning task than the High SD *Ss*. Our study showed that the reverse of this prediction was found for the female groups and that we could reach no conclusion with regard to the male groups. The Taylor and Spence formulation thus led to results which are inconsistent with predictions based on their approach.

TABLE 5

Means, Sigmas, and *t* Tests for Immediate Recall

(N = 10 per group)

Group	Mean No. of Items	Sigma	*t* Tests
High SD males	11.1	1.91	0.49
Low SD males	11.6	2.63	
High SD females	12.4	2.55	2.74*
Low SD females	9.4	2.32	

* Significant at the .05 level.

TABLE 6

Means, Sigmas, and *t* Tests for Delayed Recall

(N = 10 per group)

Group	Mean No. of Items	Sigma	*t* Tests
High SD males	10.2	2.53	0.63
Low SD males	9.4	3.13	
High SD females	11.1	3.32	2.54*
Low SD females	7.9	2.18	

* Significant at the .05 level.

The Edwards formulation was accurate in predicting the females performance, but failed to account for the males performance. It is possible that a clue to the explanation for this descrepancy occurs in the suggestion of Heilizer, Axelrod, and Cowen (1954). They obtained a significant second-order interaction "anxiety level × sex × experimenter" which forced them to calculate analyses of variance separately for each sex. They found that for the men alone, a significant difference in means appeared between the three groups tested by the three experimenters. Though this finding was not pertinent to their study, they felt it was worthy of mention. It suggested that such obvious variables as the sex of the *S* and the sex ×

experimenter interactions may be important.

The "anxiety level × sex × experimenter" interaction would very nicely account for the findings of this study. Unfortunately, the fact that only one male experimenter did all the testing rules out the possibility of evaluating this interaction. *If* this interaction were in operation during the learning procedure, then high drive in the Edwards sense, measured by a High SD score, combined with the presence of an authority figure of the opposite sex, would set up the necessary conditions for the operation of this factor, and account for the results we obtained. It could also be predicted that if a female authority figure did the testing, the results would have reversed, with the High SD males outdoing the Low SD males, while the female groups would not differ between themselves to any significant degree. It would be of interest to repeat this type of study using male and female experimenters, arranging the experimental design so as to evaluate the interaction of drive level × *Ss*' sex × experimenter.

Another possibility that could be posited to explain the results is that the SD scale really measured the self confidence or level of aspiration of the *Ss*, and that this underlying factor was related to the learning performance and accounted for the significantly superior learning and recall of the High SD females. Two arguments oppose this interpretation. If it were true, then it would follow that the High SD males would be similarly motivated and perform better than the Low SD males. This was not the case since no significant difference in performance was found between the two male groups. Secondly, a nonsignificant correlation between the self-confidence score and actual trials to mastery was obtained, indicating this factor could not account for a significant amount of the variance obtained in the learning scores.

The finding that the High S-V items were learned in fewer trials than the Low S-V items may be explained on the basis of faulty experimental methodology. Three uncontrolled sources of error could be posited to account for the results. First, the *E* read the nonsense syllables aloud during the learning procedure. He could have given subtle vocal emphasis to nonsense syllables connected with the High S-V items, and less vocal emphasis to the Low S-V items. Second, the items were recalled aloud by the *S*, and it was possible that the *E* allowed slightly more time for recall of the High S-V items than for the Low S-V items. The *E* used a stop watch to time his intervals which is far from satisfactory when equal time intervals are desired. Related to this same problem was the fact that the statements varied in length and therefore required different lengths of time merely to say them aloud. Thirdly, it was possible that by use of expressive mannerisms of which he was unaware, the *E* may have communicated the idea to the *S* that he favored the High S-V items, thereby encouraging the *Ss* to greater effort toward learning the High S-V items.

Correction of these sources of error is possible. Statements of equal length could be developed. A memory drum to present each item for an equal interval of time could be used to eliminate any tendency to favor one item type over another. Also, an auditory signal could be used to time the five-second interval of recall. These mechanical conveniences would leave the *E* free to attend to the recording of data and any other duties involved in the experimental procedure. Lastly, the *E* could sit out of the line of vision of the *S* but still be able to observe his behavior.

Summary and Conclusions

This study attempted to study the relationship between social desirability as a psychological variable and learning performance and recall of complex verbal material, using college students as subjects

It was found that social desirability, as a property of the items, accounted to a significant degree, for the superiority of the High S-V items over the Low S-V items in learning, but not in recall.

The social desirability variable, as a property of the person, was found to discriminate to a significant degree, good from poor performers in the female college population, though it failed to separate the male students on the basis of their learning performance.

Explanations of the findings were offered and alternative interpretations of the results discussed.

REFERENCES

EDWARDS, A. L. The relationship between judged desirability of a trait and the probability that the trait will be endorsed. *J. Appl. Psychol.*, 1953, **37**, 90–93. (b)

EDWARDS, A. L. *Manual for the Edwards Personal Preference Schedule.* New York: Psychological Corp., 1953. (a)

EDWARDS, A. L. *The social desirability variable in personality assessment and research.* New York: Dryden, 1957.

FORDYCE, W. Social desirability in the MMPI. *J. consult. Psychol.*, 1956, **20**, 171–175.

HEILIZER, F., AXELROD, H. S., & COWEN, E. L. The correlates of manifest anxiety in paired-associates learning. *J. Pers.*, 1954, **24**, 463–474.

LAZARUS, R. S., DEESE, J., & HAMILTON, R. Anxiety and stress in learning: The role of intra-serial duplication. *J. exp. Psychol.*, 1954, **47**, 111–114.

LUCAS, J. D. The interactive effects of anxiety, failure, and interserial duplication. *Amer. J. Psychol.*, 1952, **65**, 59–66.

RAMOND, C. K. Anxiety and task as determiners of verbal performance. *J. exp. Psychol.*, 1953, **46**, 120–124.

TAYLOR, JANET A. The relation of anxiety to conditioned eyelid response. *J. exp. Psychol.*, 1951, **41**, 81–92.

TAYLOR, JANET A., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61–64.

WOODWORTH, R. S. *Experimental psychology.* New York: Holt, 1938.

(Received December 8, 1958)

Journal of Educational Psychology
Vol. 51, No. 2, 1960

RELATIONSHIPS AMONG SEVERAL MEASURES OF PUPIL ADJUSTMENT[1]

IRA J. SEMLER
University of Texas

Currently the employment of paper-and-pencil personality, adjustment, and rating scales with school children is widespread and increasing. While sophisticated investigators and test constructors recognize no one-to-one relationship between test results and actual classroom adjustment, there exists a less critical attitude on the part of many educators. Frequently naive interpretations of group test results lead to gross discrepancies in judging an individual pupil's adjustment. A common error is failure to recognize that adjustment is a function of many variables and that different tests may measure different aspects of adjustment. An unwarranted assumption is often made that pupils who are classified at one level of adjustment by one test are similarly classified by other types of measures. While several investigations have attempted to determine relationships between certain adjustment measures and pupil characteristics (Bedoian, 1953; Bonney, 1943; Grossmann & Wrighter, 1948; Scandrette, 1953; Sewell & Haller, 1956), Tindall (1955) has pointed out that few studies have used a variety of approaches to the measurement of adjustment. An investigation which demonstrates patterns of interrelationships between certain adjustment instruments and measurement technics would also provide information for more effective interpretation of test results.

The purpose of this investigation was to determine patterns of interrelationships among three separate approaches to the measurement of adjustment; namely, peer acceptance, self-report, and authority figure (teacher) ratings. Additionally, sex differences in relationship patterns and generality of relationships across classrooms was studied.

[1] Based on a portion of the data gathered for a dissertation at the State University of Iowa.

Method

Subjects

The *Ss* were 247 boys and 236 girls, some 483 children who were enrolled in 18 fifth-grade classrooms in the public elementary school system of a Midwestern city of about 80,000 population.

Measures

Two measures of *peer acceptance* were employed. The first instrument used a sociometric (SOC) procedure patterned after that described by Jennings (1948). The second was the Ohio Social Acceptance Scale (OSA).[2] This scale provides every *S* an opportunity to rate every other *S* in his group on a friendship scale which ranges from 1 (My very, very best friends) to 6 (Dislike them).

Self-descriptions (self-report) were total scores obtained from the California Test of Personality (CTP), Elementary Series, Form AA (California Test Bureau, 1953).

Teacher ratings (TRt) of pupil adjustment were obtained on Items I, II, VI, VII, and XI of the Rating Scale For Pupil Adjustment (Science Research Associates, 1953). Ratings from the first four items were pooled to provide an adjustment score, while Item XI was used as a "school conduct" rating.

In addition to the foregoing measures of adjustment, data regarding intelligence, achievement, and parental occupation were obtained from the child's cumulative record. Intelligence quotients were based on the Otis Quick Scoring Mental Ability

[2] Ohio State Department of Education, The Ohio Scholarship Tests (undated).

Tests; the median grade equivalent earned on the Stanford Achievement Test was used as a general indicator of achievement level. Parental occupation was rated, using a 21-point modification of the Warner Index of Status Characteristics Occupation Scale.

Procedure

Teacher ratings from the SRA Rating Scale For Pupil Adjustment were secured along with cumulative record data before classroom testing in order to prevent rater contamination from knowledge of other adjustment measure results.

A planned schedule for two testing sessions one week apart enabled classroom testing for *all* 18 classrooms to be completed within a two-week period. Absentees were tested within three days after the completion of classroom testing. The same procedure was followed in all classrooms. The first session of 35 min. was used to administer the SOC and to start the CTP; the second session of 35 min. was used to administer the OSA and to complete the CTP. This was adequate time for all *Ss* to complete the two peer acceptance instruments and the self-report.

Results and Discussion

The major results of this investigation are summarized in Table 1. The pattern of obtained correlations was similar for girls, boys, and combined sexes and reflected a positive and reliable relationship between each of the pupil adjustment measures. The results in Table 1 indicate that the various instruments and measurement technics do not classify pupil adjustment in the same way even though the several approaches appeared to contain a common element which could be labelled "adjustment." The two peer acceptance instruments were highly related, which is not surprising since they are operationally similar. The next highest relationships were obtained between the two peer acceptance measures and the authority figure ratings, which suggests a common "social reaction" and an awareness on the part of teachers that social acceptability is an important aspect of adjustment. The lowest correlations were obtained between the "social" (SOC, OSA, and TRt) approaches and the self-report. Since all correlations are reliably different from zero, the conclusion is that pupils who were more acceptable socially were rated as better adjusted by their teachers and rated themselves higher (well adjusted) on a self-report.

TABLE 1

Relationships among Several Measures of Pupil Adjustment: Intercorrelations by Sex for Total Groups

	SOC	OSA	TRt
OSA			
Girls	.67		
Boys	.70		
Total	.68		
TRt			
Girls	.49	.54	
Boys	.40	.53	
Total	.45	.54	
CTP			
Girls	.30	.35	.46
Boys	.27	.37	.27*
Total	.28	.37	.37

Note.—SOC is sociometric; OSA is Ohio Social Acceptance Scale; TRt is Rating Scale For Pupil Adjustment; CTP is California Test of Personality. Top value is correlation based on all girls. Middle value is correlation based on all boys. Bottom value is correlation based on sexes combined. All correlations are reliably different from zero at the .01 level of confidence.

* In this one instance the correlations for the two sexes differs at the .05 level of confidence.

The pattern of relationships was similar for girls and boys, with the exception of a significantly higher relationship between teacher ratings of girls' adjustment with girls' self-reports (CTP) than between teacher ratings of boys' adjustment with the boys' self-reports. Since 15 of the 18 teachers were women, a further separation of relationship data (median correlations) seemed worthwhile and is presented in Table 2 according to sex of teachers as well as sex of pupils. The significant sex differences in correlations noted above is diffi-

TABLE 2

MEDIAN CORRELATIONS BETWEEN TEACHER RATINGS OF ADJUSTMENT AND OTHER MEASURES OF ADJUSTMENT ACCORDING TO SEX OF TEACHER AND SEX OF PUPIL

	SOC	OSA	CTP
Female teacher ratings of adjustment (TRt) (N = 15)			
Girls	.55**	.66**	.45*
Boys	.38	.65**	.34
Male teacher ratings of adjustment (TRt) (N = 3)			
Girls	.35	.39	.42
Boys	.68**	.68**	.46*

Note.—Significance levels based on minimum class size of N = 21.

* Reliably different from zero at the .05 level of confidence.

** Reliably different from zero at the .01 level of confidence.

cult to explain since correlation patterns involving teacher ratings of same and opposite sex pupils' adjustment and the pupils' score on a self-report were quite alike. However, Table 2 clearly shows that relationships involving teacher ratings and sociometric status differed depending upon the sex of the teacher and the sex of the pupil. The relationship between adjustment ratings (TRt) and peer acceptance (SOC) was higher for girls than for boys when female teachers were ratings, and the reverse was true when male teachers rated adjustment. All correlations involving teacher ratings of same sex pupils' adjustment were reliably different from zero at the .05 level of confidence, whereas only one correlation involving teacher ratings of opposite sex pupils' adjustment was significantly different from zero.

It is reasonable to suppose that each classroom was a unique social entity and that the total correlations reported in Table 1 did not reflect accurately the relationships for each classroom. Table 3 summarizes the results of the analyses of covariance which tested the homogeneity of the 18 classroom regression coefficients for all combinations of pupil adjustment measures. The analyses support the use of general regression lines for combinations of the sociometric with teacher ratings (SOC-TRt) and with the self-report (SOC-CTP). Significant F tests for all other pupil adjustment combinations indicated heterogeneity of classroom regression data. It follows that separate relationship data specific to the classroom should be used when combinations of these adjustment measures are considered.

Table 4 was prepared to summarize certain individual classroom statistics. It is quite noticeable that the median correlation was a poor indicator of average relationship, since the range of classroom correlations was quite wide in every case. This wide range of correlations involving teacher ratings might be accounted for partially by individual teacher difference in judging adjustment. For instance, correlations between teacher ratings of adjustment and teacher ratings of school conduct ranged from .21 to .87 (not tabled), which indicated that some teachers tended to equate adjustment with "school conduct." Of course, the low correlations within this range might mean that some teachers recognized that well behaved pupils are not always well adjusted and that a little nonconformity is not necessarily symptomatic of poor adjustment.

TABLE 3

RESULTS FROM SUMMARIES OF ANALYSES OF COVARIANCE FOR VARIOUS COMBINATIONS OF PUPIL ADJUSTMENT MEASURES: F TESTS OF HOMOGENEITY OF 18 CLASSROOM REGRESSION COEFFICIENTS

	SOC	OSA	TRt
OSA	1.94*		
TRt	1.14	5.56**	
CTP	.77	4.60**	2.97**

* Significant at less than the .05 level of confidence.

** Significant at less than the .01 level of confidence.

Despite the wide range of relationship data across classrooms, inspection of the 18 intercorrelation matrices revealed strikingly similar patterns which followed the pattern of median correlations in Table 4. In almost every classroom the correlations within the "social judgment" approaches were higher than between these approaches and the self-report. The generality and applicability of this finding should be explored further.

Table 5 summarizes the relationships between the various measures of pupil adjustment and pupil characteristics used in this study. Little difference was noted between median and total correlations.

While the linear correlations reported in Table 5 appeared to show that pupil characteristics were poor "predictors" of pupil adjustment, it is possible that relationships between some of these measures might be described better by curvilinear correlations. For example, a scatterplot of results from the Ohio Social Acceptance Scale versus median grade equivalents from the Stanford Achievement Test revealed a curvilinear trend. The computed eta for this scatterplot was .71 in comparison with obtained linear correlations of .34 (total) and .39 (median). The trend indicated that underachievers were not as socially acceptable as achievers or overachievers. Inspection of the scatterplot also showed greater variation in peer status for overachievers than for underachievers. The prediction of curvilinear trends between measures of intelligence or achievement and other measures of pupil adjustment seems reasonable and might well apply to other elementary school populations.

The effective use of group test results requires that interpreters be aware not only of extent of relationship among different approaches to the measurement of adjustment but, also, that such relationships may be relatively unique to individual classrooms. In addition, the interpreter should give especial consideration to teacher ratings, since sex of teacher with respect to

TABLE 4

MEDIAN CORRELATION AND RANGE OF CORRELATIONS FOR 18 CLASSROOMS

	SOC	OSA	TRt
OSA			
Median	.72**		
Range	.48 to .84		
No. Sig.	(18)		
TRt			
Median	.50*	.62**	
Range	.23 to .65	.25 to .84	
No. Sig.	(15)	(14)	
CTP			
Median	.30	.42	.39
Range	.05 to .49	.19 to .63	−.07 to .68
No. Sig.	(4)	(11)	(10)

Note.—Top value is median correlation of 18 classrooms. Middle values are extremes of correlation range. Bottom value, in parenthesis, is the number of correlations in the range which were reliably different from zero at the .05 level of confidence.

* Reliably different from zero at the .05 level of confidence for $N = 21$.

** Reliably different from zero at the .01 level of confidence for $N = 21$.

TABLE 5

SUMMARY OF RELATIONSHIPS BETWEEN PUPIL ADJUSTMENT MEASURES AND ABILITY, SCHOOL ACHIEVEMENT, AND PARENTAL OCCUPATION

	SOC	OSA	TRt	CTP
Intelligence (Otis)				
Median	.29	.35	.34	.36
Total	.22**	.32**	.32**	.32**
Achievement (Stanford)				
Median	.22	.39	.48*	.38
Total	.22**	.34**	.40**	.40**
Social class (Parental occupation)				
Median	.26	.28	.20	.14
Total	.22**	.23**	.26**	.20**

Note.—Median refers to median correlation of 18 classrooms. Total refers to correlation of total groups without regard to classroom.

* Reliably different from zero at the .05 level of confidence.

** Reliably different from zero at the .01 level of confidence.

sex of pupil may be an important factor when assessing the individual pupil's adjustment.

Summary

This investigation obtained measures of relationship among several approaches to the measurement of pupil adjustment. Cumulative record information (intelligence, achievement, and parental occupation) plus teacher ratings of pupil adjustment for 483 fifth-grade children in 18 classrooms were obtained in advance of a carefully planned classroom testing schedule. The classroom testing schedule obtained successive measurements of peer acceptance (SOC and OSA) and pupil responses on a self-report test (CTP).

The findings in brief were:

1. A reliable and positive but *moderate* set of intercorrelations among the several pupil adjustment measures was obtained.

2. Significant sex differences were evident in correlations involving teacher ratings and self-reports.

3. Sex differences in relationship patterns depended upon sex of teacher when teacher ratings and sociometric status correlations were considered.

4. General regression data did not describe individual classroom relationships.

5. One peer acceptance instrument (OSA) appeared to be related to certain pupil characteristics in a curvilinear fashion.

REFERENCES

Bedoian, V. H. Mental health analysis of socially over-accepted, socially under-accepted, overage and underage pupils in the sixth grade. *J. educ. Psychol.*, 1953, **44**, 366–371.

Bonney, M. E. Personality traits of socially successful and socially unsuccessful children. *J. educ. Psychol.*, 1943, **34**, 449–472.

Grossmann, Beverly, & Wrighter, Joyce. The relationship between selection-rejection and intelligence, social status, and personality amongst sixth grade children. *Sociometry*, 1948, **11**, 346–355.

Jennings, Helen H. *Sociometry in group relations*. Washington, D. C.: American Council on Education, 1948.

Michigan Department of Mental Health. *Rating scale for pupil adjustment*. Chicago: Science Research Associates, Inc., 1953.

Scandrette, O. C. Classroom choice status as related to scores on components of the California Test of Personality. *J. educ. Res.*, 1953, **47**, 291–296.

Sewell, W. H., & Haller, A. O. Social status and the personality adjustment of the child. *Sociometry*, 1956, **19**, 114–125.

Thorpe, L. P., Clark, W. W., & Tiegs, E. W. Californa test of personality. Los Angeles: California Test Bureau, 1953.

Tindall, R. H. Relationships among indices of adjustment status. *Educ. psychol. Measmt.*, 1955, **15**, 152–162.

(Received August 9, 1958)

Journal of Educational Psychology
Vol. 51, No. 2, 1960

THE EFFECT OF TEACHER–PUPIL CONTACTS INVOLVING PRAISE ON THE SOCIOMETRIC CHOICES OF STUDENTS[1]

NED A. FLANDERS AND SULO HAVUMAKI
University of Minnesota

Many teachers have used sociometric data to assign a choice value to any individual in a group. The individual who receives the most choices has the highest choice value, and the rest of the group can be arranged in rank order. Very little has been said about how teachers can influence the choice value of a student.

The present study is an attempt to see if teacher-pupil contacts involving praise will affect the sociometric choices received later on. The hypothesis to be tested is that contacts involving praise, given by a prestige figure, will increase the choice value of the student.

The following conditions must be controlled to test this hypothesis:

1. The group situation must involve an attractive activity and the praise given by the prestige figure must be centered on the relationship between an individual's behavior and the major activity of the group; merely to talk about the individual per se would not be a fair test.

2. Information given to the group must be controlled by the teacher so that certain individuals can be singled out for attention.

3. Any acquaintanceship patterns existing in the group prior to the experimental discussions must be identified so that it can be demonstrated that the factor of personal acquaintance did not bias the results.

4. The criterion question used to elicit sociometric choices must be related to the major activity of the group and appear to have consequences in the future activities of the group. This eliminates such questions as, "Who contributed the most to this discussion?"

5. The prestige figure controlling the flow of information must be liked by the group, or at least not disliked.

PROCEDURE

In a study conducted at the University of Minnesota, Laboratory for Research in Social Relations, 33 groups of 10 subjects each (hereafter referred to as *Ss*) were involved in an experiment in which sociometric choices were made. In all groups the *Ss* consisted of approximately an equal number of boys and girls from tenth-grade classes of Minneapolis and St. Paul high schools. It was the belief of these groups that they were to appear in a Quiz Kid contest between schools from Minneapolis and St. Paul. Each group was to decide whether their group would appear on a radio or TV quiz program, either alternative being possible. This discussion was led by a teacher-trainer, not known to any of the students, whose function it was to help the students reach a decision and then train them for the performance. Elaborate arrangements were made to impress all *Ss* that the Quiz Kid contest was an attractive and exciting experience. Practically all *Ss* approached the problem of choosing between radio and TV programs seriously. At the very end of each experimental session when all *Ss* learned there would be no contest, sincere expressions of disappointment were dis-

[1] The authors wish to express their appreciation to the Minneapolis and St. Paul school officials and students whose cooperation was essential to the success of the study. This research was carried out as part of the research program of the Laboratory for Research in Social Relations, University of Minnesota.

played in each group. Judging from the manner in which all sessions were conducted and the behavior of the *Ss*, there is no doubt in the minds of the authors that the prospect of participating in the contest was attractive and stimulating to the *Ss*. Thus, these group discussions satisfactorily meet the first half of Condition 1 mentioned previously.

Besides the attractiveness of the group's activity, Conditions 1 and 2 refer to the nature of the teacher-trainer and the method of giving this information. The teacher-trainer's behavior for 17 groups was individually oriented, that is, he called on certain individuals by name and praised their contributions to the discussion. The teacher behavior thus served to provide praise for certain individuals in the group by focusing attention on these individuals in the individually oriented condition. The individuals attended to in this manner were always those who happened to sit in the odd numbered seats around a U shaped table. The original choice of seats was up to the *Ss*. The procedure of calling only on *Ss* seated in the odd numbered seats effectively restricted all verbal communication to these *Ss* and the teacher-trainer. While many *Ss* seated in the even numbered seats raised their hands indicating a desire to speak, they were not called on. Thus the teacher-trainer's praise reactions were clearly centered on *Ss* who were easily identified during the individual approach.

In the remaining 16 groups the teacher-trainer behaved in a group oriented fashion and praised the group as a whole for the ideas contributed by individual members. He called on no individuals by name and anyone who wished to talk was allowed to do so.

The verbal behavior of the teacher-trainer was categorized from voice recordings of all the group discussions. These categories consisted of a simple analysis of praise statements made by the teacher-trainer in response to student contributions. In the individually oriented condition most praise statements were made with reference to specific individuals; in the group oriented condition the praise statements were never coupled with an individual's name and were interpreted in terms of the group whenever this could reasonably be done.

Praise statements became standardized during pilot runs prior to conducting the discussions reported here. Examples of statements used in the individual approach are: "I'm glad John mentioned that point because it is very important for this group to consider in order to make a good decision"; "That's a very good suggestion, Mary"; "That's very true. I think Jim will be a big help to this team." During a group approach it was not always easy to turn a praise statement away from the individual toward the group, for example: "That's an important point. This team has good ideas"; "Good! A second good point; can anyone [looking at the group] add any more?" Occasionally, during the group approach only, the teacher-trainer would be interrupted after saying, "That's a good idea..." before praise could be turned toward the group. Thus, some praise statements during the group approach were classified as individual praise statement. An example of a praise statement that is clearly group oriented is: "This group is going to make a fine team."

The comparison of the praise statements of the teacher-trainer's individual and group oriented behavior patterns shown in Table 1 indicates that neither pattern is 100% consistent. Actually, the differences are greater than indicated in Table 1. For example, the data do not reveal such important aspects of the teacher-trainer's behavior as looking directly at an individual, calling on him by name, and praising his ideas in association with his name, as contrasted with looking generally at the whole group, not calling on any one individual, and never asso-

ciating the name of an individual with praise. The two patterns of teacher behavior were clearly different to an observer and easily identified by the tape recording with 100% accuracy.

Thus far the discussion has dealt with the first two conditions necessary to the study; next, Conditions 3 and 4 will be considered. After each group was seated, they were first asked to list the names of all the *Ss* with whom they were acquainted and, in addition, place an "X" beside the name of any person who was a good friend. For all experimental groups the indication of friendship was so infrequent (a total of six friendship pairs) that they are not reported here. This was undoubtedly due to the cooperation of school authorities who were asked to choose *Ss* for each group from separate homerooms. The pre-discussion acquaintanceship data show that 477 *Ss* were acquainted with the 85 *Ss* who sat in the odd numbered seats before the discussion started; similarly, 505 *Ss* were acquainted with those *Ss* who sat in the even numbered seats. Of the total (928) acquaintanceships reported, 48.6% involved *Ss* in the odd numbered seats, 51.4% involved *Ss* in the even numbered seats. The hypothesis that this difference is significant can be tested by chi square. Since chi square = 0.012, the difference is clearly not significant. In addition, the *Ss* seated in the even numbered seats are slightly better known, so that a more rigorous test of the main hypothesis is possible.

After each discussion the teacher-trainer collected sociometric information in response to the statement, "List five members of the group whom you think would be good program participants." It was explained that two teams, size 10 each, would result in 20 *Ss* being in the studio ready to answer the quiz questions, and this was far too many. Therefore, each team would have five program participants who would answer the questions and five helpers who could help the participants. The *Ss* were told that their own group choices would be used to identify the program participants. The results of their selection are shown in Table 2.

Results

The data in Table 2 follow the prediction that *Ss* seated in the odd numbered

TABLE 1

Distribution of Teacher Praise Statements

	Individual Condition (17 groups)		Group Condition (14 groups)[a]	
	N	%	*N*	%
Praise to Individuals	114	79.1	28	30.1
Praise to Group	30	20.9	65	69.9
Totals	144	100.0	93	100.0

[a] Two groups lost as recordings were inadvertently erased.

TABLE 2

Distribution of Choices for Program Participants

	Individual Treatment (17 groups)		Group Treatment (16 groups)	
	N	%	*N*	%
To Odd Seats	496	59.3	384	48.6
To Even Seats	341	40.7	406	51.4
Totals	837*	100.0	790**	100.0

Note.—Since some *Ss* made only four choices, * does not equal 5 × 17 = 850 and ** does not equal 5 × 16 = 800.

TABLE 3

Desire to Continue with Same Teacher-Trainer

	Individual Treatment		Group Treatment	
	N	%	*N*	%
Yes	141	82.9	136	85.5
Not Sure	27	15.9	20	12.6
No	2	1.2	3	1.9
Totals	170	100.0	159[a]	100.0

[a] One *S* failed to answer this question.

seats in the individual treatment and given praise by the teacher-trainer would receive more choices than would occur by chance, and the same comparison in the group treatment would show no such trend. The significance of 59.3% distribution compared with a hypothetical 50–50 split can be calculated by using the standard error of correlated proportions. Testing the significance of a 59.3–40.7 split by this method yields a critical ratio of 2.98, indicating a 0.0028 probability that this difference could have occurred by chance. If one ignores the probable interdependence of the choice data and uses a chi square test on the data in Table 2, chi square = 19.063, indicating a probability of 0.001 for one degree of freedom.

The fifth condition listed at the beginning of this article refers to the requirement that the *Ss* like or at least not dislike the teacher-trainer. After the discussions, the teacher-trainer would leave, and the students were informed by the host-experimenter that it was not necessary for them to continue with the teacher-trainer who had led the discussions unless they wanted to. They were instructed to indicate their feelings toward the teacher-trainer by checking the item, "Would you like to continue working with the same teacher-trainer?"

Table 3 indicates that there is little difference between the two conditions of teacher behavior as to the degree of liking or of disliking of the teacher-trainer. The data for 329 *Ss* are shown in Table 3.

The hypothesis that the distribution of answers for the individual treatment are not significantly different from the distribution of answers in the group treatment can be tested by chi square. A chi square = 0.966 for two degrees of freedom fails to reject this hypothesis.

Summary

In a decision making experiment sociometric data were collected from 330 tenth-grade students after they had interacted with a teacher who praised their participation. In 17 10-man groups the teacher interacted only with *Ss* seated in the odd numbered seats. In another 16 groups, all *Ss* were allowed to talk and statements involving praise were directed to the group as a whole. In the former situation the *Ss* in the odd numbered seats received significantly more choices than did the *Ss* seated in the even numbered seats. In the latter situation, the difference between *Ss* in the odd and even numbered seats was not significant.

There are undoubtedly many different ways to increase the choice value of students in a classroom. The results of this experiment indicate that teacher-pupil interaction involving praise that is supportive and constructive is likely to increase the choice value of a student indicating greater acceptance by his peers.

(Received August 28, 1959)

Journal of Educational Psychology
Vol. 51, No. 2, 1960

THE JOPLIN PLAN OF READING VS. A TRADITIONAL METHOD

ELMER F. MORGAN, JR.
Monroe County Schools

AND GERALD R. STUCKER
Dundee Elementary Schools

The research described here was designed to determine whether the Joplin plan of reading (Floyd, 1954) would produce significantly superior results where teachers have been randomly assigned to experimental and control groups, and where sex, method, IQ, and initial reading ability have been carefully controlled.

As a rather explicit approximation of the Joplin plan is described in the Procedure, a description will not be undertaken here. It should be noted, however, that since Floyd's report the Joplin plan has been carefully reviewed and adapted by a substantial number of school systems. The plan has achieved national recognition through its popularization in a 1957 issue of the *Saturday Evening Post*. While many of the school systems using this plan claim superior results at the middle grades, there appears to be no published statistical research to validate these claims.

Subjects and Procedure

Table 1 presents the sex, IQ, and initial reading characteristics of 90 matched pairs of fifth- and sixth-grade *Ss* of a rural consolidated school located in a southeastern Michigan county. The *Ss* of Table 1 were those *Ss* out of a total population of 226 fifth and sixth graders who could be successfully matched on two separate measures of initial reading ability. The two measures of reading ability were the Durrell-Sullivan, Form A, and California, 1957 W Series, reading achievement tests.

At each grade the *Ss* were divided into fast and slow experimental and control groups by the following procedure. Those *Ss* scoring above the expected grade norms of 4.8 and 5.8 for next year's fifth and sixth grades (groups were formed at the end of the 4th and 5th grades) were designated fast achievers. Those *Ss* scoring below this cutting line were designated slow achievers. *Ss* at the fast levels were matched on the two separate reading measures and were then randomly assigned to either a control or experimental group. The procedure was repeated at the slow levels. This procedure resulted in matched experimental and control groups of two levels at each separate grade.

While the *Ss* were not matched across the rows on an intelligence test as they were on the two reading measures, the *t* test for matched groups revealed that the above groups could not be viewed as being other than from a common population also with respect to California (57 S-Form) IQs.

The eight prospective teachers of the above groups agreed to strict randomization with respect to teaching fast and slow experimental and control groups.[1] Randomization was carried out by drawing class assignments from a hat.

The plan in operation called for teachers of the control groups to teach reading as they always had in their tra-

[1] The writers wish to thank the fifth- and sixth-grade teachers of the Dundee, Michigan, elementary schools whose keen desires for experimentation made this study not only possible, but mandatory. Especially helpful was Harry Nelson, Superintendent, who gave both financial and administrative assistance.

ditional self-contained classrooms. The experimental groups separated at a set hour, and each *S* went to a reading class of his or her own level. A reading period of 50 minutes was set for both the experimental and control groups. This period was carefully controlled for all groups and reading for all groups was carried out at approximately the same morning hour. The plan was in operation for the entire 1958–1959 academic year.[2]

Limitations of Study

Because this was a single-school experiment, several limitations that were inherent in the nature of the design could not be escaped.

1. The factor of control groups meant that the time of four teachers was used up for these groups. This seriously limited the number of experimental reading groups of different levels that could be formed. For this reason, advanced reading groups beyond the intact above average sixth-grade group could not be formed. Thus, with the exception of having five highly gifted fifth graders enter the group, this group was not truly an experimental group.

2. The use of two levels of initial reading achievement meant that the experimental groups were not matched against truly heterogeneous control groups, but instead were matched against groups that were already somewhat homogeneous with respect to reading.

3. Because of the limitation imposed by the use of control groups, a considerable body of experimental *Ss* received reading instruction as well as all other instruction from the same experimental teacher. On the surface, it would appear that there was no difference in methods for these *Ss*. This limitation, however, is not as severe as it might at first appear, as the nucleus of *Ss* remaining in their homerooms were still experimental *Ss*, all of which were now receiving instruction in classes where reading range had been reduced by several grades. At the fast fifth-grade level, for example, the placing of 13 *Ss* into reading classes outside their homeroom reduced the reading range from 4.8–9.2 to 5.5–7.2. The *Ss* coming into this group from the slow sixth grade were all within the 5.5 to 7.2 reading range. At the other levels, a similar reduction in reading range was obtained. It was possible, however, to check on the above limitations by testing the null hypothesis that there was no difference in reading results for those experimental *Ss* leaving their homerooms for reading instruction and their matched

TABLE 1

Sex Composition, Mean IQ, Reading Means, and Standard Deviations of Fast and Slow Experimental and Control Groups

Group	Sex and IQ			California		Durrell-Sullivan	
	M	F	IQ	Mean	*SD*	Mean	*SD*
5th S-E	13	7	96.45	4.17	.55	3.90	.50
S-C	14	6	97.30	4.29	.39	3.83	.50
F-E	14	13	110.81	5.48	.87	5.47	.71
F-C	10	17	109.00	5.67	.79	5.51	.71
T-E	27	20		4.92	.99	4.77	.95
T-C	24	23		5.08	.94	4.76	.97
6th S-E	11	5	97.86	5.41	.72	4.78	.66
S-C	11	5	98.01	5.39	.78	4.70	.61
F-E	11	16	116.93	7.46	.92	7.30	.76
F-C	16	11	117.67	7.49	.93	7.25	.71
T-E	22	21		6.70	1.31	6.36	1.42
T-C	27	16		6.70	1.34	6.30	1.40

Note.—S-E = Slow Experimental; S-C = Slow Control; F-E = Fast Experimental; F-C = Fast Control; T-E = Total Experimental; T-C = Total Control.

[2] During the summer of 1958, book carts and supplementary reading materials were purchased for all eight classes. A new and unfamiliar basal reader series was purchased for the experimental classes. All participating teachers had equal access to a well stocked storeroom of supplementary readers, reading materials, and hundreds of library books from the county library. The participating teachers asked that no one teacher be given any supervision, as they felt that teaching conditions should remain the same for all. None was given throughout the entire year.

counterparts in the control classes. The same hypothesis was stated for those experimental *Ss* remaining in their homerooms for reading instruction. Should significant results be obtained in favor of the experimental groups, failure to reject the null hypothesis of equal effectiveness of method for those experimental *Ss* leaving their homerooms would tend to lessen the meaning of any significant trends.

4. When the above limitations are taken into consideration, it can only be concluded that this experiment is not a fair representation of the Joplin plan at its best. Because of these limitations, trends significant at less than the .10 level will be taken as trends indicating the worthwhileness of further experimentation.

Results

At all levels, the *t* test for matched groups and the Wilcoxon matched-pairs signed ranks test (Siegel, 1956), a nonparametric test, were used to test the null hypothesis that the two different teaching methods produced no significant difference in reading results. Since the number of *Ss* was small, and since there was no ground to support the assumption of normality, the nonparametric test will be regarded as the more critical test at the fast and slow levels.

Table 2 presents the reading means and levels of significance by both the *t* and Wilcoxon tests on an alternate form of the Durrell-Sullivan. For both the fifth and sixth grades the results are reasonably clear-cut. The reading means of the fast and slow experimental groups are significantly higher at well less than the .10 level, with the exception of the fast sixth-grade level where none was expected (see Limitation 1). When the fast and slow levels are combined to form total experimental and control groups, the difference is significant at less than the .01 level at the fifth grade and at less than the .05 level at the sixth grade.

Table 3 presents the same analysis as Table 2 on an alternate form of the California. Here it may be seen that results are in line with the results obtained on the Durrell-Sullivan. It is interesting to note that the means of the California are significantly higher than those of the Durrell-Sullivan at both the fast and slow levels. Since it was not the reading measures that were initially matched, but rather the *Ss* themselves on the two independent measures, this finding has only academic interest.

Turning next to Limitation 3, we find that the 23 fifth-grade *Ss* leaving their homerooms for reading instruction made significantly higher mean reading scores on both reading measures than did their matched counterparts in the control classes. The means for the Durrell-Sullivan were 5.68 to 5.40, and 6.33 to 5.90 for the California. The null hypothesis of equal effectiveness of teaching method was rejected at less than the .05 level by the Wilcoxon test for both reading measures. The null hypothesis was also rejected at less than the .05 level in favor of the remaining 24 experimental *Ss* who received

TABLE 2

Durrell-Sullivan Reading Means and Tests of Significance by the *t* and Wilcoxon Tests

Group	Durrell-Sullivan		*t* Ratios			Wilcoxon Test	
	Number	Mean	$\frac{\bar{d}}{SE_{\bar{d}}}$	*t*	*P*	*T*	*P*
5th S-E	20	4.93	.190	2.00	< .10	51	< .05
S-C	20	4.55					
F-E	27	6.72	.126	2.46	< .05	78	< .05
F-C	27	6.41					
T-E	47	5.95	.119	2.77	< .01		
T-C	47	5.62					
6th S-E	16	5.83	.220	3.59	< .01	14	< .01
S-C	16	5.04					
F-E	27	8.04	.221	.76	—	136	—
F-C	27	7.87					
T-E	43	7.22	.165	2.24	< .05		
T-C	43	6.85					

TABLE 3

CALIFORNIA READING MEANS AND TESTS OF SIGNIFICANCE BY THE t AND WILCOXON TESTS

Group	California		t Ratios			Wilcoxon Test	
	Number	Mean	$S\bar{s}_1 - \bar{s}_2$	t	P	T	P
5th S-E	20	5.83	.181	2.65	<.02	39	<.02
S-C	20	5.35					
F-E	27	7.18	.142	2.04	<.06	80	<.05
F-C	27	6.89					
T-E	47	6.60	.130	2.85	<.01		
T-C	47	6.23					
6th S-E	16	6.58	.203	1.82	<.10	31.5	<.07
S-C	16	6.21					
F-E	27	8.50	.120	1.66	>.10	121	>.10
F-C	27	8.30					
T-E	43	7.78	.110	2.36	<.05		
T-C	43	7.52					

reading instruction in their homerooms. At the sixth-grade level only 14 *Ss* left their homerooms for reading instruction (see Limitation 1). Of these 14 experimental *Ss*, 11 made higher reading scores on both tests, which called for a rejection of the null hypothesis at less than the .05 level by the Wilcoxon test. Thus, with respect to Limitation 3, it can be concluded that the experimental plan of reading was approximately as effective for those *Ss* leaving their homerooms for reading instruction as it was for those *Ss* remaining in their homerooms.

DISCUSSION

Attempts to explain the obtained superiority of the experimental plan are of course at this time highly conjectural. Nevertheless, we would like to suggest the following reasoning. The placing of students into reading groups where the reading range and variance have been greatly reduced means that the individual student can safely and securely face up to his reading problem without losing face. For the slow student, he is now functioning in a nonthreatening group of other problem readers who are experiencing problems similar to his. It is our hypothesis that this nonthreatening situation allows the student to receive maximum positive reward or feedback from reading material *he can now read*. This rarely happens in heterogeneous competitive groups at Grades 4, 5, and 6 where grouping within the class is practically nonexistent. Such grouping also allows the teacher to develop better techniques to handle the restricted reading range and thereby maximizes her opportunities to offer verbal and emotional rewards. Once again, however, it is our hypothesis that the nonthreatening group atmosphere which allows for a maximum positive feedback from readable materials is the more important factor of the two. The advantages for the high achievers are so obvious that they need not be discussed.

The results obtained do appear to offer rather clear support for the notion that the Joplin plan is a more effective plan of reading than the traditional plan of self-contained classrooms with reading ranges of from six to eight grades at Grades 4, 5, and 6. In actuality, this suggests that teaching or more probably learning effectiveness increases as the group becomes more truly homogeneous.

Concerning this experiment, care must be taken not to generalize the significant findings to other schools. Further experi-

mentation may well show that the plan is not equally effective for all schools. Should other investigators obtain results similar to these, especially in multischool designs, an intensive analysis or observation of these groups at work should prove productive in isolating some of the teaching and learning variables that make for more efficient learning, something we still know little about, despite a host of academic learning experiments.

Summary

At the conclusion of Grades 4 and 5, 180 of 226 *Ss* were matched separately on the Durrell-Sullivan and California reading achievement tests. The *Ss* of each grade scoring above and below the expected grade norms of 4.8 and 5.8 were designated fast and slow achievers. The fast and slow *Ss* of each grade were then randomly assigned to experimental and control groups, which resulted in matched control and experimental groups on two levels. Experimental and control groups at all levels were also found to be from a common population on the California Test of Mental Maturity.

Teacher assignments to these groups were carried out by random assignment. Supplementary reading materials and teacher supervision were equalized. The experimental *Ss* received reading instruction according to the Joplin plan. The control *Ss* received reading instruction in traditional self-contained classrooms. The experiment ran for one year, at which time the experimental groups at all levels made significant gains. It was concluded that, for the single school where the experiment was run, the Joplin plan was more effective. It was suggested that the obtained superiority might be due to the lowered variance of the experimental groups, which allowed the teacher to offer more effective verbal and emotional rewards, but more particularly allowed the slow student to function in a nonthreatening atmosphere which maximized positive feedback from readable materials.

REFERENCES

Floyd, C. Meeting children's reading needs in the middle grades: A preliminary report. *Elementary Sch. J.*, 1954, **55**, 99–103.

Seigel, S. *Nonparametric statistics*. New York: McGraw Hill, 1956.

(Received August 29, 1959)

Journal of Educational Psychology
Vol. 51, No. 2, 1960

A NOTE ON THE TEACHING OF SKINNER'S DESCRIPTIVE BEHAVIORISM IN THE FIFTH GRADE[1]

H. E. KLUGH, W. A. DETERLINE
Alma College

AND D. E. HENDERSON
East Jordan Public Schools

The work of Ojemann, Levitt, Lyle, and Whitside (1955), Stiles (1950), and others has demonstrated that teaching elementary school children a "dynamic" or "causal" approach to behavior can be successfully accomplished with appropriately trained teachers and materials. Psychology taught in the elementary grades thus far has had, primarily, a mental health emphasis. In this paper we will report the results of a rather informal preliminary program for the inclusion of a section on psychology as an integral part of the elementary school natural science course.

Procedure

The third author (Henderson), who did the teaching described, had one semester's previous experience teaching elementary school, and had completed a minor in psychology as an undergraduate. One of the courses making up this minor was Analysis of Behavior which used *Principles of Psychology* by Keller and Schoenfeld (1950), and included a three-hour weekly laboratory in which each pair of students demonstrated principles of learning using a white rat, a Skinner box, and a laboratory manual by Homme and Klaus (1957). Henderson, after listening to some of the questions his elementary students posed concerning animal behavior, decided that a section might be added to his natural science program incorporating many of the experimental studies described in his college manual.

The experimental class consisted of 33 unselected fifth-grade children. The mean California IQ was 100. A 12-item pretest covering the proposed material was administered to the experimental and to a control class of 34 children whose mean California IQ was 104. The means of the 12-item pretest indicated no significant pretreatment differences.

The method of procedure was to begin the class period with a brief discussion of the phenomena to be demonstrated. The children then grouped around the teacher's desk on which was a Skinner box equipped with a manually operated water reinforcing device. Students acted as timers and recorders, and manipulated light S^Ds and reinforcement. In meetings of about 30-min. duration twice a week over an eight-week period the instructor discussed and demonstrated reinforcement, extinction, spontaneous recovery, discrimination, generalization, the concept of S^D and S^Δ and the development of operant chains. At the conclusion of each demonstration the data were recorded on a blackboard, and all children produced graphs showing the appropriate functional relationships. This was the first use of graphs for these children. Emphasis throughout the demonstrations was on careful observation of the animal's behavior and its possible relationship to both controlled and uncontrolled external conditions. The children noted, somewhat in advance of the instructor, the

[1] Presented at American Psychological Association meeting, Cincinnati, September, 1959.

effect of a stop watch click acting as a possible secondary reinforcer. At the end of each demonstration the discussion period attempted to show how the principle which operated with the rat might be used to predict and explain aspects of the behavior of pets, the children themselves, or parents and teachers. These discussion sessions were recorded on tape and sent to the college where suggestions and comments were added and the tapes returned. Midway through the program a chapter of reading material (Klugh, 1959) was circulated to the children, and this was apparently of some assistance.

Results and Comment

The results of a readministration of the 12-item achievement test indicated a significant increase in knowledge of and verbal application of concepts for the experimental class ($t = 3.2$, $df = 60$, $p = .01$). The control group posttest showed a slight decline in mean test score from the pretest condition. It is difficult to evaluate such a program. The authors were impressed by the great enthusiasm the children showed for the course, the care with which they made their observations, and the ease with which they could ask incredibly difficult questions. The authors feel that with appropriately trained teachers, apparatus, and reading material there is reason to believe that rudimentary understanding of the simpler functional relationships in behavior can be taught in the science program of the later elementary grades and that such a program will be enthusiastically received by the children.

Summary

An unselected fifth-grade class of 33 children served as subjects in an exploratory study of the limits and possibilities of teaching simple concepts of Skinner's behaviorism in the elementary grades. Skinner box demonstrations were given twice a week as part of the regular science program. Concepts demonstrated were reinforcement, extinction, secondary reinforcement, discrimination, generalization, partial reinforcement, and chaining. Children operated the apparatus, drew graphs, and took part in discussions relating the concepts demonstrated to human behavior. The children were enthusiastic, interested, and apparently acquired an elementary understanding of these concepts in terms of human behavior, as measured by a brief achievement test.

REFERENCES

Homme, L. E., & Klaus, D. J. *Laboratory studies in the analysis of behavior.* Pittsburgh: Lever Press, 1957.

Keller, F. S., & Schoenfeld, W. N. *Principles of psychology.* New York: Appleton Century, 1950.

Klugh, H. E. Behavioral science for the later elementary grades. Unpublished manuscript, Alma College Library, Alma, Michigan, 1959.

Ojemann, R. H., Levitt, E. E., Lyle, W. H., Jr., & Whitside, Maxine F. The effects of a "causal" teacher-training program and certain curricular changes on grade school children. *J. exp. Educ.*, 1955, **24**, 95–114.

Stiles, F. S. Developing an understanding of human behavior at the elementary school level. *J. educ. Res.*, 1950, **43**, 516–524.

(Received October 22, 1959)

Journal of Educational Psychology
Vol. 51, No. 2, 1960

ATTITUDE CHANGES OF STUDENT TEACHERS AND THE VALIDITY OF THE MINNESOTA TEACHER ATTITUDE INVENTORY

OWEN SCOTT
College of Education, University of Georgia

AND STERLING G. BRINKLEY
Emory University

This study investigated the validity, predictive and concurrent, of the Minnesota Teacher Attitude Inventory (MTAI) for student teaching. Data obtained in the validity study were also used to investigate changes in the attitudes of student teachers during student teaching.

Subjects and Procedure

Eighty-two student teachers enrolled in seven white teacher education institutions in Georgia participated. These student teachers were enrolled for student teaching on a full-time basis for approximately 13 weeks. Data were also obtained from the supervising teachers of these student teachers and from 1,973 public school pupils. Participation was voluntary with teachers and pupils in Grades 4 through 12 included. Two-thirds of both the teachers and the pupils were in Grade 9 and above. Subjects taught by the high school teachers included English, mathematics, science, social science, and home economics.

The study of attitude changes employed three sets of data: MTAI scores of student teachers obtained within 10 days prior to student teaching, MTAI scores of their supervising teachers obtained early in the student teaching period, and MTAI scores of student teachers within two or three days after completion of student teaching. For the predictive and concurrent validity studies, these MTAI scores were used, and, in addition, scores obtained on a paper-and-pencil measure of student-teacher-pupil personal relations were utilized as criterion measures.

Descriptions of teacher behavior and of the relationship of teacher behavior to MTAI score contained in the MTAI manual indicate that teachers scoring high on MTAI tend to establish classroom personal relations customarily described as "democratic," while low scoring teachers tend to establish relations usually characterized as "undemocratic" (Cook, Leeds, & Callis, 1951, pp. 3–4). The selection or construction of a criterion instrument for measuring "democratic personal relations in the classroom" posed quite a problem. After a careful study of relevant literature and of many instruments, the investigators decided to describe "democratic personal relations in the classroom" and to devise an instrument based on this description (Bryan & Yntema, 1939; Grim & Hoyt, 1952; Ryans, 1952; Ryans & Wandt, 1952). For purposes of this study the phrase means that the student teacher is successful in (*a*) creating a friendly classroom atmosphere, (*b*) establishing in pupils (and possessing himself) a feeling of psychological security, (*c*) exerting a stabilizing influence on pupil behavior, (*d*) developing pupil self-responsibility, (*e*) helping pupils understand and accept social controls, (*f*) understanding and accepting each pupil as a person, (*g*) helping each pupil establish a feeling of his own personal worth, and (*h*) treating pupils in consistent fashion from day to day.

Using this reference frame, the investigators constructed a 15-item device to be answered by pupils, designated as the Classroom Personal Relations booklet (CPR). The first 14 items are concerned specifically with personal relations in the classroom, while the 15th is a "global" item inquiring about the pupil's estimate of the

student teacher's over-all teaching effectiveness in comparison with other teachers the pupil has had. For each item the five possible responses are designed to lie at approximately equal intervals along a continuum from "effective" to "ineffective" in terms of the aspect of personal relations dealt with. The Likert method of scoring using arbitrary weights was adopted, using weights of 1, 2, 3, 4, and 5 (Likert, 1932). Seven of the first 14 items were randomly selected to be scored in reverse with Option 1 scored as 5, etc. The arithmetic mean CPR score of a class taught by the student teacher was selected as the criterion measure.

In addition to using the reference frame referred to above in devising items, the investigators had each item carefully checked by selected elementary and high school teachers, by a number of specialists in elementary and secondary education, and by some pupils. Precautions taken to make probable the validity of CPR were at least as adequate as those taken with other instruments used in making validity studies of MTAI (Cook & Leeds, 1947; Leeds, 1950; Sandgren & Schmidt, 1956).

The equivalent halves method of estimating the coefficient of internal consistency was employed, using the first 14 CPR items. The split-half r was .72; the reliability coefficient, using the Spearman-Brown formula, was .84 (McNemar, 1955, pp. 156–157). Standard deviations of the distributions of CPR scores were found to be between 3 and 4 (range of possible scores, 14–70), so that the standard errors of measurement were in the neighborhood of 1.6, and the standard error of measurement of the mean, approximately 0.3.

At each of the seven teacher education institutions a project coordinator, usually the college supervisor working with the supervising and student teachers, was designated to obtain voluntary participants, to administer MTAI both to student teachers and to supervising teachers, to score the CPR responses, and to tabulate the data on specially prepared data sheets. Supervising teachers administered CPR after the student teachers had completed student teaching and left their schools. In the instances of student teachers who taught more than one class of pupils, either the class in which the student teacher had done the most teaching was selected, or, if the student teacher had taught about the same length of time in several classes, one class was randomly selected to respond to CPR.

Changes in the Attitudes of Student teachers During Student Teaching

Of many factors possibly associated with observed changes in attitude (e.g., influence of prior professional courses, factors traceable to the home environments of the student teachers, etc.) only one factor, the attitude of the supervising teachers, was considered in this study. Data were available for 77 student teachers. The mean attitude change of this group was determined and tested, the mean attitude changes of two sets derived from this group were similarly examined, and investigation was also made of individual changes in attitude of the student teachers. Table 1 contains data pertinent to these analyses.

All three of the student pretest-supervisor mean differences in Table 1 are significant. However, only one of the students' pre- and posttest mean differences is significant—that for the 47 student teachers whose initial MTAI scores were in each instance lower than those of their supervising teachers. Of the six r's in the last two rows, three are significant with only one being significant at the .01 confidence level. These statistics seem to show some slight degree of association between attitude changes of student teachers and the attitudes of their supervising teachers, with the association limited to instances in which the student teachers made lower initial scores than did their supervising teachers.

Analysis of changes in attitude of in-

TABLE 1

CHANGES DURING STUDENT TEACHING IN MEAN MTAI SCORES OF 77 STUDENT TEACHERS AND OF TWO SETS DERIVED FROM THE GROUP

	Students with Initital Scores		
	Lower than Supervisors (N = 47)	Higher than Supervisors (N = 30)	Total Group (N = 77)
Initial MTAI scores of students			
Mean	22.8	56.0	35.7
Standard error of mean	2.8	3.4	2.8
Poststudent teaching MTAI scores			
Mean	34.7	53.9	42.2
Standard error of mean	3.5	4.4	2.9
MTAI scores of supervising teachers			
Mean	62.7	34.3	51.6
Standard error of mean	3.1	3.1	1.7
Differences: Means of students' pre- and Posttests (MTAI)	11.9**	−2.1	6.5
Differences: Means of students' Pretests and supervisors' MTAI scores	−39.9**	21.7**	−15.9**
Correlation: Students' pre- and Posttests (MTAI)	.60**	.60**	.60**
Correlation: Students' pretests with supervisors' MTAI	.27	.27	.27*
Correlation: Students' posttest with supervisors' MTAI	.34*	.34	.34**

Note.—Correlations are Pearson r's; linearity of regression was tested in each instance.
* Significant at .05 level of confidence.
** Significant at .01 level of confidence.

dividual student teachers corroborates the association just mentioned. Nineteen of the 77 student teachers made statistically significant gains in score (.05 level of confidence), 12 underwent significant losses, and 46 underwent no significant change. Fourteen of the 19 student teachers making significant gains were in the set of 47, while 8 of the 12 undergoing significant losses were in the set of 30. The chi square (8.08 with 2 DF) computed for these data is significant beyond the .05 level of confidence, the major contribution to it being the number of significant positive changes in attitude in the set of 47. The contingency coefficient (N of 77) turns out to be .31. For the 31 students who underwent significant changes in attitude during student teaching, the r between their attitude changes and the differences between their initial scores and the scores of their supervising teachers is .42, significant at the .01 level of confidence.

PREDICTIVE AND CONCURRENT VALIDITIES OF THE MTAI

The predictive validity study focused on the question, "To what extent will the attitudes of student teachers toward pupils prior to student teaching (measured by MTAI) predict the personal relations with pupils they actually establish during student teaching?" The concurrent validity study was concerned with the question, "To what extent are attitudes toward pupils expressed by student teachers during student teaching (measured by MTAI) accompanied by the establishment of corresponding personal relations with pupils?" Answering these two questions necessitated the construction of the criterion device previously described.

One of the statistics used to determine both the predictive validity and the concurrent validity of MTAI was the partial correlation coefficient indicating the re-

lationship between student teachers' MTAI scores and the criterion measure, with supervising teacher MTAI scores held constant. For the predictive validity study the student teachers' MTAI scores just prior to student teaching were used; for the study of concurrent validity, the student teachers' MTAI scores immediately after student teaching were employed. For both studies, mean class scores on the first 14 CPR items were used as criterion measures. Table 2 presents a summary of data relevant to these two validity investigations. Data contained in Table 2 clearly indicate that neither the predictive validity coefficient nor the concurrent validity coefficient was significantly different from zero at the .05 confidence level.

Although MTAI had neither predictive nor concurrent validity in terms of mean responses to the first CPR items taken as a whole, possibly it had validity in terms of responses to individual CPR items. In both the predictive and validity study, item analyses were made for each CPR item. The 10 student teachers scoring highest on MTAI and the 10 scoring lowest were selected in each study, and CPR item responses were analyzed to determine which items differentiated significantly between the two groups of student teachers. In the predictive validity item analyses, the responses to each item of the 252 pupils in the classes of the 10 highest scoring (on MTAI) student teachers were treated initially as one distribution, and the responses of the 263 pupils in the classes of the 10 lowest scoring student teachers were similarly considered. Item analyses in the concurrent validity study, using the responses to each item of the 242 pupils in the classes of the 10 highest scoring (on MTAI) student teachers and of the 237 pupils in the classes of the 10 lowest scoring student teachers, were made in like fashion. In the analyses, responses scored as 5, 4, or 3 were considered "favorable"; those scored as 2 or 1, as "unfavorable." The percentage of favorable responses in each distribution was computed, and the significance (.01 level of confidence) of the difference between percentages was tested (McNemar, 1955, pp. 60–61). Table 3 lists for both the predictive and concurrent validity studies the items and percentages for which the difference in percentages of favorable responses was significant.

TABLE 2

PREDICTIVE AND CONCURRENT VALIDITY COEFFICIENTS (AND OTHER RELEVANT FACTS) OBTAINED FOR THE MTAI

	Predictive ($N = 74$)	Concurrent ($N = 69$)
Student teachers' MTAI scores		
Means	35.2	42.2
Standard Deviations	34.2	34.4
Supervising teachers' MTAI scores		
Means	51.6	51.6
Standard Deviations	32.7	32.7
Pupils' CPR scores		
Means	58.7	58.8
Standard Deviations	3.0	3.0
Correlations: Student teachers' and supervisors' MTAI scores	.27	.34
Correlations: Student teachers' MTAI and pupils' mean CPR scores	.03	−.04
Correlations: Supervisors' MTAI and pupils' mean CPR scores	.01	−.01
Partial Correlation: Student teachers' MTAI and pupils' CPR scores with supervisors' MTAI held constant	.03	−.04

Note.—For each r linearity of regression was tested.

In the predictive validity item analyses some items were found to discriminate positively; others, negatively. For each of these six items, 70% or more of the pupils in both groups responded favorably so that discriminations were in terms of differences in sizes of the majorities of favorable responses. The items which discriminated positively ask specifically about the interpersonal relations between pupil and stu-

TABLE 3

CRITERION ITEMS FOR WHICH DIFFERENCES IN PERCENTAGE OF FAVORABLE RESPONSES WERE SIGNIFICANT

	Predictive*		Concurrent**	
	P_U	P_L	P_U	P_L
1. How do you feel toward your student teacher?	79	70	76	60
2. How does your student teacher feel toward you?	83	72	85	77
4. Are your classmates good friends with you?	71	82	—	—
8. How do you feel about disagreeing with your student teacher?	74	85	79	67
13. Does your student teacher treat you in about the same way every day?	80	92	83	74
14. Does your student teacher understand you and your problems?	93	76	93	71
15. Does your student teacher teach as well as other teacher you have had?	—	—	82	70

Note.—For the predictive validity study, MTAI scores were obtained prior to student teaching.

For the concurrent validity study, these were obtained immediately after student teaching.

N_U : The number of pupils in the classes (one per student teacher) of the 10 student teachers scoring highest on MTAI.

N_L : The number of pupils in the classes (one per student teacher) of the 10 student teachers scoring lowest on MTAI.

P_U : The percentage of pupils responding favorably to the indicated CPR item in the classes of the 10 student teachers scoring highest on MTAI.

P_L : The percentage of pupils responding favorably to the indicated CPR item in the classes of the 10 student teachers scoring lowest on MTAI.

* In the predictive validity study N_U was 252 and N_L, 263.

** In the concurrent validity study N_U was 242 and N_L, 237.

dent teacher; the ones which discriminated negatively ask questions about the friendship of fellow pupils, about feelings concerning disagreeing with the student teacher, and about the consistency with which the student teacher treats the pupil.

Six items were also found to discriminate significantly, positively in each instance, in the concurrent validity study. As in the predictive validity study, all discriminations were in terms of differences in sizes of the majorities of favorable responses. Inspection of the content of these six discriminating items makes clear the fact that they pertain to three of the eight characteristics of "democratic personal relations in the classroom" (defined in this study): (*a*) creating a friendly atmosphere, (*b*) treating the pupil in consistent fashion from day to day, and (*c*) understanding and accepting the pupil as a person. Also involved (CPR Item 15) was the pupil's estimate of the student teacher's over-all teaching effectiveness.

Each of the two sets of responses to an item considered as one distribution in testing the significance of the difference in percentages of favorable responses was, in fact, a composite of 10 distributions. In both the predictive and concurrent validity studies, each of these composites (24 in all) was tested for homogeneity of variance and for significance of the differences among means. Homogeneity of variance among the distributions comprising each composite was tested, using Bartlett's chi square and the significance of differences among means within each composite was tested, using an F test (Edwards, 1950, Chaps. 9 and 10). In each instance in the predictive validity study in which the variances were found to be heterogeneous, two independent transformations of the scores were effected (first, a square root transformation; then, a logarithmic transformation) because of the possible effects of heterogeneity of variance upon the differences among means (Edwards, 1950, pp. 197–203). In no instance did score transformation change the results of the analysis. Since the concurrent validity data were similar to those of the predictive validity study, no transformations were made in the concurrent validity study. In only one pair of composite distributions (from CPR Item 4 in the predictive validity study) were differences among means found to be nonsignificant.

One may conclude from these data that, except for CPR Item 4 in the predictive validity study, both among the classes of

student teachers scoring highest on MTAI and those of student teachers scoring lowest, there were significant differences in responses to those CPR items which discriminated between the group of highest scoring (MTAI) student teachers and the group of lowest scoring ones. In other words, intragroup differences were significant as well as intergroup differences.

Summary

Eighty-two student teachers enrolled in seven white teacher education institutions in Georgia responded to the Minnesota Teacher Attitude Inventory (MTAI) both just prior to and just after a quarter of full-time off campus student teaching. Analysis of differences in prestudent teaching and poststudent teaching responses to MTAI were made, both for changes in group means and for changes in individual scores. The group of student teachers working with supervising teachers whose attitudes toward pupils were, in each instance, superior to their own, improved significantly as a group in their attitudes toward pupils during student teaching (i.e., the change in the group's mean score during student teaching was significant). The group working with supervising teachers whose attitudes were, in each instance, inferior to theirs, did not as a group change significantly in their attitudes toward pupils during student teaching. Further analyses, based on the attitude changes of individual student teachers, corroborated these findings.

A purpose of the study was to examine the predictive and concurrent validities of MTAI, using mean pupil responses to a carefully constructed check list, the Classroom Personal Relations booklet (CPR) as criteria. With these criteria, MTAI had neither predictive nor concurrent validity, a finding in conflict with some other studies (e.g., Hoyt & Cook, 1959; Sandgren & Schmidt, 1956; Stein & Hardy, 1957; Stogdill & Coons, 1957). When responses to individual CPR items were used as criteria, in some instances there were statistically significant differences between student teachers scoring highest on MTAI and those scoring lowest; however, these differences were of little practical significance.

References

Bryan, R. C., & Yntema, O. A manual on the evaluation of student reactions in the secondary school. Kalamazoo, Mich.: Western State College, 1939. (Mimeo.)

Cook, W. W., & Leeds, C. H. Measuring the teaching personality. *Educ. psych. Measmt.*, 1947, **7**, 399–410.

Cook, W. W., Leeds, C. H., & Callis, R. *Minnesota Teacher Attitude Inventory*. New York: Psychological Corp., 1951.

Edwards, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.

Grim, P. R., & Hoyt, C. J. Appraisal of teaching competency. *Educ. res. Bull.*, 1952, **31**, No. 4.

Hoyt, C. J., & Cook, W. W. The predictive validity of the Minnesota Teacher Attitude Inventory based on pupil attitudes toward school. *J. tech. Educ.*, 1959, **1**, 42–44.

Leeds, C. H. A scale for measuring teacher-pupil attitudes and teacher-pupil rapport. *Psychol. Monogr.*, 1950, **64** (6, Whole No. 312).

Likert, R. A technique for the measurement of attitudes. *Arch. Psychol.*, 1932, **22** (Whole No. 140).

McNemar, Q. *Psychological statistics*. (2nd ed.) New York: Wiley, 1955.

Ryans, D. G. A study of criterion data. (A factor analysis of teacher behavior in the elementary school.) *Educ. psychol. Measmt.*, 1952, **12**, 333–344.

Ryans, D. G., & Wandt, E. A factor analysis of observed teacher behavior in the secondary school: A study of criterion data, *Educ. psychol. Measmt.*, 1952, **4**, 574–586.

Sandgren, D. L., & Schmidt, L. G. Does practice teaching change attitudes toward teaching? *J. educ. Res.*, 1956, **49**, 673–680.

Stein, H. L., & Hardy, J. Validation of the Minnesota Teacher Attitude Inventory in Manitoba. *J. educ. Res.*, 1957, **50**, 321–338.

Stogdill, R. M., & Coons, A. E. (Eds.) Leadership behavior: Its description and measurement. *Ohio St. Univer. Bur. Bus. Res. Monogr.*, 1957, No. 88.

(Received January 17, 1959)

Journal of Educational Psychology
Vol. 51, No. 2, 1960

A STUDY OF MEDIATION IN MENTALLY RETARDED AND NORMAL SCHOOL CHILDREN[1]

GERSHON BERKSON AND GORDON N. CANTOR
George Peabody College for Teachers

Numerous experiments have provided evidence bearing on the mediation hypothesis, a notion which is playing an increasingly prominent role in S-R theory. One experimental paradigm found in such studies utilizes three lists of stimulu which, for present purposes, may be labeled A, B, and C. If the subject (*S*) learns paired-associates involving Lists A and B, and then ones involving Lists B and C, mediation theory predicts that the subsequent learning of paired-associates involving Lists A and C will be facilitated due to the prior learning experience. Data relevant to this formulation have been reported on college students (Peters, 1935; Bugelski & Scharlock, 1952; Russell & Storms, 1955) and on normal preschoolers (Norcross & Spiker, 1958).

The present investigation involved an attempt to demonstrate the occurrence of mediation in normal and mentally retarded elementary school children, using the A-B, B-C, A-C paradigm. An additional purpose of the study was to compare normal and retarded children in their ability to learn paired-associates and to perform a mediation task.

The experimental predictions were as follows: (*a*) the normal *Ss* would learn lists of paired-associates more quickly and with fewer errors than would the retarded *Ss*; (*b*) both the retarded and normal *Ss* would be capable of mediation behavior such as that described above; and (*c*) a difference in IQ level would not affect the degree of facilitation achieved through the mediation process.

The first prediction was based on the observation that the learning of retarded individuals is generally less efficient than that of normals. With respect to the second prediction, it was assumed that the retarded as well as the normal *Ss* could reach a learning criterion on the A-B and B-C pairs. Employing this assumption, S-R theory would lead one to expect facilitation in the learning of the A-C list in both groups of *Ss*. Again, assuming the attainment of a common learning criterion on the A-B and B-C lists for both groups, there would be no basis in S-R theory for expecting a *differential amount* of facilitation in the groups (i.e., with equal A-B and B-C habit strengths built up in both groups, one would anticipate an equal degree of facilitation on the A-C list); this served as the rationale for the third prediction.[2]

Assuming all three predictions were confirmed by the results, one would have a basis for contending that the teacher must train retarded pupils longer to reach a given level of verbal learning, but once that

[1] This investigation was partially supported by the National Institute of Mental Health through the mental deficiency training program at George Peabody College for Teachers. Acknowledgment is made to the Nashville, Tennessee City School System for permission to carry out the experiment in the Nashville public schools.

[2] An assumption is made here that the groups acquire comparable habit strengths on the A-B and B-C lists by virtue of reaching the same learning criterion. Strictly speaking, if the retarded *Ss* require more trials to reach criterion and if habit strength is defined solely in terms of number of reinforcements, it would follow that the retarded *Ss* approach the A-C list with *greater* habit strength for at least some of the A-B and B-C components. In this case, comparable facilitation on the A-C list could be attributed to such a habit differential. Because of the results actually obtained on the A-B and B-C lists, however, this complication is largely eliminated in interpreting the results of the present experiment.

level is achieved, mediation based on such learning will occur in the same degree in the retarded as in normals.

Method

Subjects and materials. The *Ss* were 54 children attending regular and special classes in the Nashville, Tennessee City School System. All *Ss* were Caucasian and, with one exception, were chosen from schools located in lower socioeconomic neighborhoods. The *Ss* were between 9 and 12 yrs. old. The *Ss* in the normal group had IQs between 86 and 115. The mentally retarded *Ss* had IQs between 55 and 85. The instruments used for measuring intelligence were the Kuhlmann-Finch Test (Forms III and V). There were 30 *Ss* in the normal group and 24 *Ss* in the retarded group. Within each IQ level, *Ss* were assigned at random to experimental and control conditions. A summary of the characteristics of the groups may be seen in Table 1.

No child was included in the experiment if his MA was below seven years, if he had been diagnosed as brain injured by a physician, or if he could not meet the learning criterion used in the study. Four normal *Ss* and 12 retarded *Ss* were excluded from the study on the basis of one or more of these criteria.

The materials consisted of three lists of paired stimuli. Each stimulus was pasted or painted on a 5 x 8 in. unlined filing card. The stimuli fell into four classes—the odd arabic numerals from 1 through 11, the even arabic numerals from 2 through 12, pictures of common objects, and hexagons varying in color. The common objects pictured were an automobile, a camera, an airplane, a baby, a girl, and a man. The hexagons were colored red, green, yellow, blue, purple, or orange.

There were six pairs of stimuli in each of the three lists. A pair was made up of two cards set back to back and encased in a transparent, colorless celluloid envelope.

Design. The design of the experiment is presented in Table 2. The independent variable was involved in the first list learned. In List I, the stimulus cards for the experimental *Ss* were the odd numerals (A). The controls received the even arabic numerals (X) as stimuli. The response cards for both experimental and control groups were the pictures of common objects (B).

Lists II and III were the same for the experimental and control groups. The stimuli of List II were the objects (B), and the response cards were the colors (C). In List III, the stimuli were the odd numbers (A), and the responses were colors (C).

Within Lists I and II, the pairing of stimulus and response cards was random. In List III, the numbers and colors were paired which, for the experimental group, had previously been paired with the same object. For example, if the number "three" had been paired with "man" in List I and "man" had been the stimulus for "yellow" in List II, then "three" and "yellow" were paired in List III.

The *Ss* learned each list to a criterion of five successive correct repetitions. No *S*

TABLE 1

IQ, MA, and CA of the Groups

Group	*N*	IQ		MA (mo.)		CA (mo.)	
		M	*SD*	*M*	*SD*	*M*	*SD*
Experimental normal	15	100.33	7.99	130.80	7.71	130.93	6.03
Control normal	15	97.27	8.98	130.27	9.65	134.27	6.09
Experimental retarded	12	70.25	6.98	99.17	7.48	142.00	10.95
Control retarded	12	69.33	6.64	95.42	6.83	138.17	11.52

TABLE 2

Design of the Experiment

Group	IQ Level					
	Normal			Retarded		
	List I	List II	List III	List I	List II	List III
Experimental	A-B	B-C	A-C	A-B	B-C	A-C
Control	X-B	B-C	A-C	X-B	B-C	A-C

was included in the analysis of the results who did not meet this criterion within 30 presentations of the full list on any of the three lists. The *Ss* who were not included in the analysis of the results were replaced by random selection of other available *Ss* from the same population.

Procedure. The *Ss* were led into the experimental room, which was either an empty classroom or testing room in the school which *S* attended. The *S* was asked to sit in a chair on the opposite side of the desk from *E*. The *E* then read the following instructions to *S*:

> This is a game in which I want to see how well boys and girls can learn which things go together. On each of these cards there is a number or a picture or a color.

The *S* was shown examples of the stimuli. At this point, all the odd numbers (A) were presented to the control *Ss*, and these *Ss* were asked to read them.[3] The *E* then proceeded with the instructions as follows:

> In the first part of the game, the picture will be on the back of the number, like this. When I show you a number, you will have to guess which picture is on the back. Do you understand? Now, here are some numbers and the pictures which go with them. Tell me the name of the number and the picture as I show them to you.

The *E* went through List I once with *S*. If *S* was unable to name one or more of the stimuli presented, he was not permitted to continue but was given the reward. Otherwise *E* continued with the instructions:

> Good. Now, if you learn what I want you to learn, I will give you a Hershey bar. O.K.? Can you tell me the name of the picture which goes with this number?

The paired-associates were presented in a prearranged, random order which changed with each successive presentation of the list. When *S* made his first error, *E* told him that he was working on a hard task, that everyone makes mistakes, and that he should not worry about making errors. When *S* had completed List I, *E* asked him to walk around the room while *E* rearranged the cards. Introducing List II, *E* said:

> This time, the pictures will be on the front and the colors will be on the back. Tell me the names of the pictures and the colors as I show them to you. Can you tell me which color goes with this picture?

Following List II, *S* was again asked to walk around the room. Before List III, *E* said:

> This time the numbers will be on the front and the colors on the back. Tell me the names of the numbers and the colors as I show them to you. Can you tell me which color goes with this number?

When *S* completed the final list, he was given a chocolate bar and returned to his classroom.

[3] The control *Ss* were asked to read the odd numbers in order to establish that they had prior familiarity with these stimuli.

Results

Analysis of variance was applied to the criterion scores by means of a 2 x 2 factorial design. The three measures analyzed were CA, number of trials taken to reach criterion, and number of errors in each list. The Kolmogorov-Smirnov test of deviation from a hypothesized distribution was applied for a test of normality of the data. The Hartley test was used to test the assumption of homogeneity of variance. No distributions were significantly different from normal. Evidence for heterogeneity of variance was found only in the error data of List I. In the subsequent analyses of variance, the .05 level of significance was used for all comparisons except those involved in the analysis of the errors of List I. The level of significance utilized for the errors in List I was .025 (Lindquist, 1953).

The analysis of CA showed no significant difference between the ages of the experimental and control groups, but the retarded *Ss* were significantly older than the normals. The test of interaction was not significant.

To determine whether analysis of covariance would be appropriate to control for the differences found in CA, Pearson correlations were run between CA and the trials to criterion and error scores for each list. No significant correlations were obtained, and it was decided to proceed with the analyses of variance.

Table 3 presents the results of List I. The analysis of variance revealed no significant differences in the learning of List I either for trials to criterion or number of errors. No significant interaction was observed. The results of List II show a slightly different pattern than did those of List I. The means and *SD*s of List II may also be seen in Table 3. While for the trials measure there were again no significant differences between any groups, the normal *Ss* did make significantly fewer errors in learning List II ($F = 5.61$, $df = 1$ and 50, $P < .05$). No significant interaction was found.

Tables 3 and 4 show the results of List III. In List III, the experimental *Ss* learned significantly more quickly and with fewer errors than did the control *Ss*. It may also be seen that on both measures the normals were more efficient than were the retarded *Ss*. No significant interaction was observed in the analysis of List III.

TABLE 3

RESULTS OF LISTS I, II, AND III

List	Group	Measure			
		Trials		Errors	
		M	*SD*	*M*	*SD*
I	Normal-Experimental	16.07	3.01	21.87	6.94
	Normal-Control	16.40	3.36	16.53	4.68
	Retarded-Experimental	16.58	5.82	19.58	11.45
	Retarded-Control	17.50	6.02	22.25	10.67
II	Normal-Experimental	11.07	3.37	8.47	6.56
	Normal-Control	11.67	4.67	10.60	11.15
	Retarded-Experimental	13.83	4.16	18.58	9.39
	Retarded-Control	11.92	3.58	12.75	10.21
III	Normal-Experimental	8.93	3.81	7.27	10.32
	Normal-Control	13.73	15.39	16.33	12.94
	Retarded-Experimental	11.25	5.95	8.08	9.52
	Retarded-Control	17.67	5.45	30.08	18.13

TABLE 4

ANALYSIS OF VARIANCE OF LIST III

Source	*df*	Trials		Errors	
		MS	*F*	*MS*	*F*
Experimental vs. Control	1	411.13	15.49*	2962.96	17.54*
Normal vs. Retarded	1	130.20	4.91*	707.20	4.19*
Interaction	1	8.71	0.33	557.65	3.30
Error	50	26.54		168.96	

*$P < .05$.

DISCUSSION

The first prediction—that normal children would learn paired-associates more efficiently than retarded *Ss* of the same CA—was only partially confirmed. The normals made fewer errors on List II and were superior to the retarded *Ss* in both measures on List III. The pattern of clearer differences between normals and retarded over trials suggests that the groups may have been learning to learn paired associates at different rates. It is possible that the technique used in the present experiment was too insensitive to detect differences in learning ability initially. With additional training, the advantage that the normals may have had by virtue of being better able to learn how to deal with the task may have produced the differences between the groups found on the second and third lists. These results, however, do make it possible to argue that the two groups encounter List III with comparable List I and List II habit strengths (see Footnote 2).

The results of List III show that both the normal and retarded mediation groups were significantly superior to the controls in speed and accuracy in learning the list. These results are consistent with the second

prediction and suggest a confirmation of the mediation hypothesis in the populations from which the *Ss* in this study were drawn. The failure to find a significant interaction involving the intelligence groups and the experimental-control treatments indicates, in line with the third prediction, that there is no evidence in this experiment for a relationship between IQ and degree of facilitation associated with mediation. Thus, whenever classroom learning can be fitted to the A-B, B-C, A-C paradigm, the results suggest that facilitation on the A-C material due to the A-B, B-C training will occur, and in the same degree, in retarded as well as normal children as long as the A-B, B-C material is learned equally well in both groups. Of course, these generalizations apply only to groups like those serving as *Ss* in the present experiment.

A comment regarding a flaw in the design of the experiment would seem to be in order. The conditions of the study permitted the experimental *Ss* fairly extensive experience with the odd numbers (List I) before these numbers were presented in List III. The control *Ss*, on the other hand, had gone through the list of odd numbers only once prior to being tested on List III. It might be argued that the advantage for the experimentals on List III could as easily have been due to greater familiarity with the materials as to the mediation condition.

While this interpretation is possible, it should be remembered that all *Ss* had had experience with odd numbers prior to the experiment and that no *S* was permitted to continue in the study who could not demonstrate a great degree of familiarity with all the elements. It may thus have been that all *Ss* were acquainted with the elements to such a degree that additional experience with the odd numbers did not substantially add to their familiarity. If the latter interpretation is accepted, one can conclude that a mediation process has been shown to aid in the learning of a list of paired-associates.

Summary

A group of 24 mentally retarded children and a group of 30 normal children were each divided into experimental and control groups. Each *S* learned three lists of paired-associates. The lists which the experimental *Ss* learned were arranged in an A-B, B-C, A-C paradigm. The control *Ss* learned Lists X-B, B-C, and A-C. The experimental and control conditions differed only in regard to the nature of the stimuli of the first list.

The results supported three conclusions: (*a*) the normal children learned paired-associates more efficiently than did the retarded children, (*b*) the mediation (experimental) condition facilitated the learning of List III for both retarded and normal *Ss*, and (*c*) the degree of facilitation in the learning of List III was not significantly related to the IQ level of the groups.

REFERENCES

Bugelski, B. R., & Scharlock, D. P. An experimental demonstration of unconscious mediated association. *J. exp. Psychol.*, 1952, **44**, 334–338.

Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Cambridge: Riverside Press, 1953.

Norcross, K. J., & Spiker, C. C. Effects of mediated associations on transfer in paired-associate learning. *J. exp. Psychol.*, 1958, **55**, 129–134.

Peters, N. H. Mediate association. *J. exp. Psychol.*, 1935, **18**, 20–48.

Russell, W. A., & Storms, L. H. Implicit verbal chaining in paired-associate learning. *J. exp. Psychol.*, 1955, **49**, 287–293.

(Received June 29, 1959)

Journal of Educational Psychology
Vol. 51, No. 2, 1960

COMPARABILITY OF PEABODY PICTURE VOCABULARY TEST SCORES UNDER GROUP AND INDIVIDUAL ADMINISTRATION[1]

RAYMOND C. NORRIS, JOHN V. HOTTEL, AND SADYE BROOKS
George Peabody College for Teachers

The present study was conducted in connection with the development and standardization of the Peabody Picture Vocabulary Test (PPVT). This test consists of 150 plates containing four pictures each. The pictures on any one plate all represent objects or activities or states of being. In administering the test the examiner provides a stimulus word orally as in "Show me *ball*." The subject then points to, or in some other way indicates, the picture on the plate which best illustrates the meaning of the stimulus word. The plates are arranged in ascending order of empirically determined difficulty, and the subject is tested from a basal of eight consecutive correct responses to a ceiling of six errors on eight consecutive plates. The raw score for a subject is his basal score plus the number of items answered correctly up to his ceiling. A more complete description of the instrument has been reported by Dunn (1959) in his manual.

PURPOSES

While the PPVT was designed to be used as an individual examination, the test author and his staff found it advantageous in trying out possible test items on some levels to administer the items to groups rather than individuals. Photographic slides of tentative item plates were projected onto a screen, and classroom groups of subjects marked their responses on separate answer sheets. The procedure worked so well that consideration was given to the possibility of collecting the normative data for normal children of late elementary school age in the same manner. However, there was no evidence that group testing would yield scores comparable to those obtained under individual administration. The primary purpose of the present study was to test the hypothesis that PPVT scores obtained under individual and group administrations would not differ significantly or appreciably. If this hypothesis were upheld, it was believed that the norm sample could be increased greatly while effecting considerable savings in time and expense.

Two other purposes helped determine the design used in the current study. In the preparation of the PPVT, item plates were selected in such a way that two items of nearly comparable difficulty could be obtained from each page. These items were then assigned to Forms A and B in such a way as to equate difficulty of the forms for each year level. The authors were interested in finding out how successful this equating process had been and how comparable scores on the two forms might be considered.

The third purpose for the present study was the determination of how much practice effect could be expected on a second administration of the test in alternate form. In assessing the outcomes of special training programs it seemed desirable that practice effect either be eliminated entirely or estimated very closely. In clinical evaluation the second form might be used for verification of diagnosis and knowledge of how much score discrepancy might be expected due to practice alone would be helpful.

[1]The research reported herein was performed pursuant to Contract SAE-6425, United States Office of Education, Department of Health, Education, and Welfare.
The generous cooperation of Lloyd M. Dunn, Project Director, and his project staff is gratefully acknowledged.

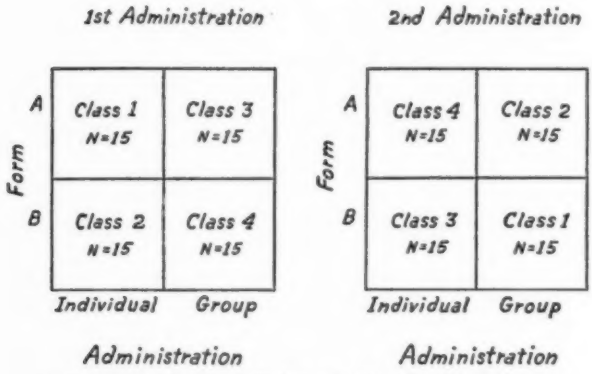

FIG. 1. Experimental design of test ad ministrations.

The three major hypotheses to be tested then were:

H_1 : PPVT scores obtained under group and individual administration do not differ significantly ($\mu_{I0} = \mu_{I1}$).

H_2 : PPVT scores based on Form A do not differ significantly from those based on Form B ($\mu_{A} = \mu_{B}$).

H_3 : PPVT scores will not be significantly altered by the subjects' having had another form of the test earlier ($\mu_{I1} = \mu_{I2}$).

METHOD

Experimental Design

The experimental design used in this study is called a Type V Mixed Design by Lindquist (1953). With two forms of the test, two types of administration, and two possible orders of administration this design made it possible for the effects of all three treatments to be evaluated as within-subject comparisons. Any given subject received both Form A and Form B, both an individual and a group administration, and both a first and second administration. The increased precision obtained by using within-subjects rather than a combination between-subjects and within-subjects variability as an error term made it more likely that any differential effects of forms, treatments, or orders would be identified.

The experimental design might be diagramed as in Fig. 1.

Subjects

The four groups of *Ss* called for in the experimental design were selected from the fifth grade of one of the Nashville City Schools.[2] Although all members of the four classes were tested, the data were analyzed only for children of appropriate age for grade placement (117 to 137 months in the second month of the school year) and of approximately normal IQ (Kuhlmann-Finch IQ scores between 80 and 121). Since it was not possible to constitute random groups for the purposes of this experiment, the selected *Ss* within each class became an experimental group. While this procedure was not considered ideal, the authors were not able to identify relevant variables such as age, sex, or IQ on which the groups differed significantly. Assignment of students to the classroom groups had been on an essentially random basis at the beginning of the year.

Preliminary screening of *Ss* on the basis of age-grade placement and intelligence left unequal numbers of *Ss* within the four subgroups. To facilitate the analysis *Ss* were

[2] The cooperation of W. H. Oliver, Superintendent of Schools, Rupert Klaus, School Psychologist, along with the principal and teachers involved in the study, is gratefully acknowledged.

eliminated at random until each group contained the same number as the smallest subgroup. In this way the number of cases per group was reduced to 15. Table 1 provides a summary of sex, IQ, and CA data on the *Ss* in each experimental group.

TABLE 1
BACKGROUND DATA ON EXPERIMENTAL GROUPS

Group	IQ		CA (Months)		Sex	
	Mean	*s*	Mean	*s*	M	F
1	97.47	11.05	124.93	4.33	6	9
2	97.07	10.86	128.53	6.05	9	6
3	101.73	9.91	125.73	4.98	7	8
4	97.73	10.35	126.13	5.01	7	8
Combined	98.50	10.19	126.33	5.18	29	31

Procedure

On the first day of the experiment six examiners administered the appropriate form of the test to individual *Ss* in the first two groups. The same six examiners also administered the appropriate form of the test to the remaining two intact groups. On the third day of the experiment the same six examiners administered the alternate form of the test to the remaining two intact groups. On the third day of the experiment the same six examiners administered the alternate form of the test to those individuals who had been tested in groups earlier. The *Ss* who were tested individually on the first day were then given the alternate form in a group presentation. Individual tests were administered to these children in the main room of the school library, a small workroom which adjoins the library, the school clinic, or the private study of the school principal. Each child was brought into a testing room and given the following instructions:

I am going to show you some pictures. There are four pictures on each page—one, two, three, four (pointing at the same time to an appropriate picture on Plate No. Example A). I will say a word, and I want you to tell me the number of the picture that best tells the meaning of the word.

The child was then presented with the three sample plates, Examples A, B, and C, and the proper form of response was established. The examiner then said:

That's fine. Now I'm going to show you some more pictures. I'll say a word, and I want you to give me the number of the picture that best tells the meaning of the word; one, two, three, or four. As we get further along, you'll find that you may not be sure of the meaning of the word. I want you to look carefully at all four pictures anyway, and make the best choice you can. Be sure to look carefully at all the pictures.

As the test progressed, each child was given verbal encouragement and admonished several times to "look at all the pictures." Testing was begun at the recommended beginning point (Plate No. 70) for each child. The plates were presented in ascending order. If a child committed an error before he had achieved eight successful identifications, the examiner turned back to Plate No. 69, and the plates were presented in descending order until a total of eight consecutive pictures were identified. Eight consecutive successful identifications constituted a basal point, and the child was assumed to be able to identify correctly any picture below that point. Once the basal point was established, testing was resumed at the point where the error was committed and proceeded until the child committed six errors out of any eight consecutive presentations (chance performance). At this point it was assumed that he would continue to perform at only chance level.

The number of the plate where testing was terminated was recorded. The total number of errors committed was subtracted from this, and the remainder was the child's raw score.

Group Test Situation. The plates were shown to each of the four classes as a group. They were transferred to 35 mm. transparencies and projected onto a 60″ X 60″ screen. One examiner operated the

projector; one called out the appropriate words (each word was given twice); and the remaining four examiners served as proctors to insure that the children were starting at the proper place and that each understood the directions. Group testing was conducted in the following manner. The projector and screen were prepared and an examiner drew a facsimile of the answer sheet on the chalk board. He then said:

I am going to show you some pictures. They will appear here on this screen. There will be four pictures on each slide; one, two, three, and four (pointing to the four appropriate positions). You see that there are three blocks like this one (pointing to illustration on chalkboard) on the half sheet of paper attached to your answer sheets. Mrs. Brooks will say a word; I want you to look at the picture on the screen and place an X in the block which shows the picture that best tells the meaning of the words she says. Let's try some.

Examples A, B, and C were then presented, and the proctors helped each child who indicated lack of understanding of, or confusion in, following the directions. When it was ascertained that all children in the class understood the procedure, the examiner said:

Good, now I'm going to show you some more pictures, and Mrs. Brooks will say the words. I want you to look carefully at all four pictures and place an X over the position of the picture that best tells the meaning of the word Mrs. Brooks says. As we get further along, you will find that you aren't sure you know the word. But I want you to look carefully at all the pictures anyway and make the best choice you can. Remember to look at all the pictures.

Testing was then begun with Plate No. 45 and proceeded through Plate No. 120. Plates No. 45 through 69 were presented for a period of 10 sec.; Plates No. 70 through 120 for 15 sec. From Plates No. 80 through 120 the examiner admonished with each fifth plate: "Be sure to look at all the pictures" or "Be sure to look at all four pictures and make the best choice you can."

Scoring of group answer sheets was begun at Plate No. 70 and a basal established in the same manner as for individual testing; that is, scoring progressed upward from No. 70 to point of first error. If eight successful identifications were made before an error was committed, scoring continued until six errors were committed out of eight consecutive presentations. If an error was committed before eight successful identifications had been made, scoring proceeded backward from point of first error to point where eight consecutive successful identifications were obtained, and this point was considered basal. No errors were counted below this point. Six errors

TABLE 2

PPVT Performance of Sixty Fifth-Grade Children under Experimental Conditions

	Form		Administration		Order	
	A	B	Individual	Group	1st	2nd
Mean	76.67	76.75	77.08	76.23	75.97	77.35
Standard Deviation	9.90	10.32	10.39	9.81	9.64	10.52

TABLE 3

Analysis of Variance

Source of Variation	*df*	Σx	s^2	*F*	$F_{.05}$	$F_{.01}$	$F_{.10}$
Within							
Subjects	60	1,490.50					
Form (F)	1	1.01	1.01	.04	4.00	2.80	1.68
Administration (A)	1	21.67	21.67	.86	4.00	2.80	1.68
Order (O)	1	57.41	57.41	2.29	4.00	2.80	1.68
F × A × O	1	5.21	5.21	.21	4.00	2.80	1.68
Error (w)	56	1,405.20	25.09				
Between							
Subjects	59	10,576.49					
F × A	1	429.41	429.41	2.44	4.00	2.80	1.68
F × O	1	143.00	143.00	.81	4.00	2.80	1.68
O × A	1	147.41	147.41	.84	4.00	2.80	1.68
Error (b)	56	9,856.67	176.01				
Total	119	12,066.99					

out of eight consecutive items established the ceiling and no credit was given for successful identification beyond this point. The raw score for each child was computed in the same manner as the raw scores for individual presentations.

No child completed the group tests without a ceiling, or terminal point, being established before Plate No. 120 was reached. It was necessary to establish basal points for six children by individual presentation. This was done on the following school day. No more than four individual plates had to be presented to establish these basal points.

Testing Times. About 10 min. were required for individual presentation, and approximately 30 min. for group testing.

Analysis and Discussion

Means and standard deviations for all 60 *Ss* on each form, under each type of administration, and for each order were determined. These data are summarized in Table 2.

From Table 3 it can be seen that form of the test, type of administration, testing order, nor any combination of these factors produced a change in PPVT scores which could not be attributed to chance at the .05 level of significance. The obtained *F* ratio for the Form A–Form B comparison was so small as to leave little doubt as to the equivalence of the two forms for the population represented by the group tested. Similarly, an *F* ratio greater than that obtained in the individual–group administration comparison would be obtained more than 50% of the time by random sampling alone; considerable confidence in the equivalance of scores obtained by individual and group methods seems warranted.

While an *F* ratio greater than that obtained between first and second administrations of the PPVT could be expected between 10 and 20% of the time by chance alone, the obtained difference of 1.38 raw score points was in the direction of a practice effect. Consequently, it might be concluded that the practice effect, if any, amounted to only about one point.

Conclusions

When both forms of the PPVT were administered in counterbalanced order to fifth-grade children of normal intelligence and in the proper school grade for their age, their average scores were not a function of the form administered nor of having the test administered individually or in a group. Furthermore, any practice effect which might have stemmed from having had the alternate form two days earlier averaged only about one point.

REFERENCES

Dunn, L. M. *Manual: Peabody Picture Vocabulary Test.* Nashville: American Guidance Service, 1959.

Lindquist, E. F. *Design and analysis of experiments in psychology and education.* New York: Houghton Mifflin, 1953.

(Received August 10, 1959)

Journal of Educational Psychology
Vol. 51, No. 2, 1960

SOURCES OF TENSION IN MARRIED AND SINGLE WOMEN TEACHERS OF DIFFERENT AGES

MARVIN POWELL
Western Reserve University

AND CHARLES D. FERRARO
National Aeronautics and Space Administration

It was the aim of this study to determine among female teachers the age differences in degree of conflict evident in various areas of psychological adjustment. The various areas of psychological adjustment examined were those which previous studies (Peck, 1936; Symonds, 1943; Barker, 1948) have indicated are of some concern to teachers and school administrators. These areas are: (*a*) Emotional Tendencies, (*b*) Heterosexual Relations, (*c*) Religion, (*d*) Vocational Outlook, (*e*) Social Acceptability, (*f*) Teacher-Supervisor Relationships, and (*g*) Physical Health.

The paucity of experimental data on adjustments of teachers precludes the making of adequate generalizations on the adjustment areas examined in this study. Actually there have been few studies of adults in general in the areas of adjustment investigated in this study (Hurlock, 1953; Harsh & Schrinckel, 1950; Young, 1947), though Havighurst (1953) has speculated regarding the type of developmental tasks which characterize various age groups and which might be expected to represent areas of adjustment.

Since married women are probably less dependent than single women on their jobs, it was hypothesized that they would be less likely to accept supervision, especially with increasing age. Many of the satisfactions of the married teachers in such areas of adjustment as Emotional Tendencies, Heterosexual Relations, Religion, Social Acceptability, and Physical Health are derived from their homes and families, and perhaps augmented by their jobs. Thus, it was felt that little disturbance should be evidenced in these areas. The area of Vocational Outlook was also considered to be nondisturbing, since it was presumed that married women were more likely to be employed in a field of their own choice.

Unmarried teachers, on the other hand, must rely completely on their jobs, and if their choice of teaching as a career was a poor one, they are "stuck" with it. Thus, it was expected that they would show disturbance in the areas of Vocational Outlook and Teacher-Supervisor Relationships. As marriage becomes less and less likely with increasing age, it was hypothesized that the areas of Heterosexual Relations, Emotional Tendencies, Social Acceptability, and Religion would show increasing disturbance, as the single woman tries to make other adjustments as a substitute for marriage. Physical Health should also be of great concern to the single woman, since her income depends on her being present to perform her job.

Generally, then, it was proposed that single women would be more disturbed in most areas at all age ranges than married women.

Procedure and Subjects

The Stimulus Material

Most studies of adjustment problems have employed check lists or questionnaires. In contrast, the present study utilized responses and reaction times in word association tests as indirect measures of psychological conflict in the above mentioned areas of adjustment. Previous research (Sharp, 1937; Powell, 1952) has

shown that reaction time in response to stimulus words, as well as the type of response given to a conflict word, differ from the reaction time and type of verbal response given to a neutral word. Thus, it seems reasonable to assume (*a*) that word symbols stand as surrogates for actual experience in a given area of adjustment, (*b*) that there will be a longer reaction time to words in the area in which the teacher is experiencing the greater degree of conflict, and (*c*) that the types of verbal response given a "conflict" word will differ from the types of verbal response given to a "neutral" word.

Stimulus words having emotional associations with each of the areas mentioned above, with the exception of Teacher-Supervisor Relationships and Physical Health, as well as "neutral" words, were selected from the word lists employed by Powell (1955). Additional words having emotional connotations in the areas of Teacher-Supervisor Relationships and Physical Health were selected from a list of stimulus words having possible emotional associations in these areas prepared with the assistance of undergraduate students in child psychology. This list of words was then checked against Rinsland's Word List (1945), and only those words appearing for fifth graders were selected to be judged by 10 Education graduate students.[1] The words finally selected to represent the two critical areas mentioned above were those among 75 surviving the above procedure on which there was 100% agreement among the 10 Educational Students as to area of adjustment represented.

In its final arrangement for experimental use, a practice series of five words was given to each *S* at the beginning of the test to clear up any misunderstandings and to familiarize the *S* with the technique to be followed. This was followed by the initial neutral word series of seven words to determine the average neutral reaction time for each *S*. Finally, each critical list of five words for the adjustment areas was followed by a neutral list of four words. These neutral words were inserted between each critical word series to act as a partial control for any "generalization" effect which has been noted in previous studies (the emotional tone aroused by the critical stimulus words extending to the subsequent reaction causing the appearance of a "disturbance" where no disturbance actually exists). However, this generalization effect was not noted in Powell's study (1952) nor in this study.

The critical word series (indicated by astericks) and the intervening neutral word series, together with the initial neutral word series and practice series were:

Practice Series: White, Cat, Man, Black, Girl
Initial Neutral Series: Canary, Counter, Ladder, Lantern, Lighthouse, Shadow, Tower

**Emotional Tendencies:* Worry, Afraid, Unhappy, Restless, Anxious
Neutral Series: Apple, Window, Minnow, Piano

**Heterosexual Relations:* Dance, Kissing, Marriage, Dates, Hugging
Neutral Series: Second, Tree, Bench, Sand

**Religion:* Church, God, Heaven, Worship, Prayer
Neutral Series: Leaves, Fender, Magazine, Sandwich

**Vocational Outlook:* Wages, Money, Employed, Hire, Job
Neutral Series: Daylight, Photograph, Bean, Monday

**Social Acceptability:* Popular, Friend, Lonely, Party, Unfriendly
Neutral Series: Cardboard, Banana, Harmonica, Door

**Teacher-Supervisor Relationships:* Superior, Authority, Supervisor, Principal, Boss
Neutral Series: Sleep, Carpet, Hill, Floor

**Physical Health:* Sick, Hospital, Invalid, Pain, Doctor

[1] Rinsland's list gives the frequency of use of words at each grade level as determined from various papers (compositions, theses, etc.) done by children at each age level.

TABLE 1

MEAN AGES IN YEARS OF EXPERIMENTAL POPULATION

(N = 25 in each group in each age range)

Age Range	Mean Age of Married Teachers	Mean Age of Unmarried Teachers
20–29	24.6	24.2
30–39	35.1	34.4
40–49	45.5	44.7
50–59	54.8	55.6

Neutral Series: Lamp, Elevator, Table, Balloon

At the conclusion of the test the *S* was asked to give information as to date of birth and marital status. The entire test was administered in privacy in a room, well lighted, containing a table or desk and chairs along with the experimental apparatus. The reaction time apparatus used in this study and the procedure followed in administering the directions were the same as those used by Powell (1952).

Subjects

A total of 200 white female public school elementary teachers (100 married and 100 unmarried)[2] ranging in age from 20 to 59 years participated in this study. All *Ss* in the investigation were teaching in elementary public schools in the Greater Cleveland area at the time the study was made, and were selected so as to comprise samples of 25 married and unmarried individuals in each of four age categories at age ranges as follows: (*a*) 20–29, (*b*) 30–39, (*c*) 40–49, and (*d*) 50–59. The mean age for each of these groups in each age range is presented in Table 1.

All *Ss* in this study were assured that they would remain anonymous and that results obtained for each individual would be kept confidential. *Ss* in this study held at least a bachelor's degree from an accredited college or university.

Methodological Considerations for Analyzing Obtained Data

The accuracy of the electric chronometer and the changes in the *E*'s speed in manipulating the switch which stopped the electric chronometer were checked on various occasions during the course of this investigation as described by Powell (1952). In addition, two independent observers were employed to check the reliability of the reaction time recorded. Each observer (*O*) used a separate switch to control his own chronometer. The reaction times recorded by the *O* and the *E* were compared for the purpose of obtaining the differences in reaction time. The results indicated a correlation of .90 between *O*'s and *E*'s reaction time.

A prime factor to be considered in reaction time (the interval between the instant the stimulus word is presented and the instant the response word is given by the *S*) is that each individual has a reaction time that is peculiar to him. A second factor to be considered is the possibility of changes in reaction time, per se, with age, which has been demonstrated in a number of studies. For these reasons it seemed best to use each *S* as his own control in order to correct for any possible changes in gross reaction time with age, and to derive a mean difference score for each individual in each area of adjustment.

The mean difference score was obtained for each individual by determining each individual's average reaction time to the initial "neutral word" list and then subtracting therefrom the individual's average reaction time to "critical" words in each of the areas of adjustment. (For example, if an individual's average reaction time to the initial "neutral" series was 1.96 seconds and his reaction time to a "critical" word was 2.76 seconds, the difference score was 0.80 seconds). The aver-

[2] For the purpose of this study, the unmarried woman is one who has not been married at any time before this study. The married woman is one who was living with her spouse at the time of this study. Thus, this investigation excludes widows, divorcees, and those otherwise separated from their spouses.

age of the five difference scores in each critical area was the mean reaction-time score used in the statistical analyses in this investigation.

Difference scores were computed for the married group and the unmarried group at each of the various age ranges. It was hypothesized that as the degree of conflict in the various areas of psychological adjustment increased there would be a slower reaction time to the "critical" stimulus words in these various areas. The Fisher-Behrens *d* test (1939) was employed to determine whether the increase in reaction time is greater than that which could be accounted for by sampling errors from a population of such difference scores.

The *d* test was used in this investigation rather than analysis of variance, since the age ranges at which significant differences in adjustment should have appeared were predicted before the collection of data for this study, and the *d* test was considered to be a precise test of the differences postulated. With this test it is not necessary to assume that the two samples come from populations having the same variance. The formula used to determine *d* in this study is the raw score formula presented by Johnson (1949), and by using the level of confidence tables derived by Sukhatme (1938).

Qualitative Word Analysis

For a qualitative analysis the words given by the *Ss* with reference to the stimulus words were classified into six categories:

Equivalent: Definition of the stimulus word, including synonyms
Logical: Any logical relation to the stimulus word, including antonyms
Personal: Any association of a personal tvpe, including names, likes, or dislikes
Perseverative: Any repetition in series. Same word repeated regardless of stimulus word
Irrelevant: No relation to the stimulus word
No response

There was 91% agreement among three judges and the *E* as to proper categorization.

In the final analysis, the verbal responses which were made by the *Ss* were placed into two categories—those considered "nondisturbed," which included equivalent plus logical and those considered "disturbed," which included personal, perseverative, irrelevant, and no responses. Any finer breakdown seemed inadvisable as reported by Powell (1952) in his attempt to categorize responses in his study.

Results and Discussion

An analysis of the mean reaction time scores to the initial neutral scores was made before attempting to test for age and marital differences in the various areas of psychological adjustment. Means and standard deviations for the several age and marital groups are contained in Table 2. A comparison of the *d* test of the reliability of the difference between married and unmarried groups at the four successive age ranges yielded *d*'s of 1.028, 1.612, 1.463, and 0.324, respectively. None of these was of sufficient magnitude to warrant rejection of the null hypothesis. Comparison of married individuals from one age group to the next (20–29 vs. 30–39; 30–39 vs. 40–49; 40–49 vs. 50–59) yielded *d*'s of 1.438, 2.918, and 0.865, re-

TABLE 2

Means and Standard Deviations of Reaction Time Scores in Seconds of Married and Unmarried Teachers at Various Age Ranges to Initial Neutral Words

(N = 25 in each group in each age range)

Age	Married		Unmarried	
	Mean	*SD*	Mean	*SD*
20–29	1.601	0.300	1.522	0.241
30–39	1.483	0.280	1.612	0.286
40–49	1.747	0.355	1.607	0.321
50–59	1.666	0.305	1.636	0.349

spectively. Only the second (30–39 vs. 40–49) was significantly large to reject the null hypothesis, and this at the .01 level of confidence. Similar *d*'s in the case of the unmarried were 1.205, 0.058, and 0.306, none of which was large enough to warrant rejection of the null hypothesis.

It is of interest that age and marital differences in reaction time were negligible, even though the "difference scores" utilized in other comparisons removed the influence of such differences in the comparisons to be turned to next.

Mean reaction time difference scores to

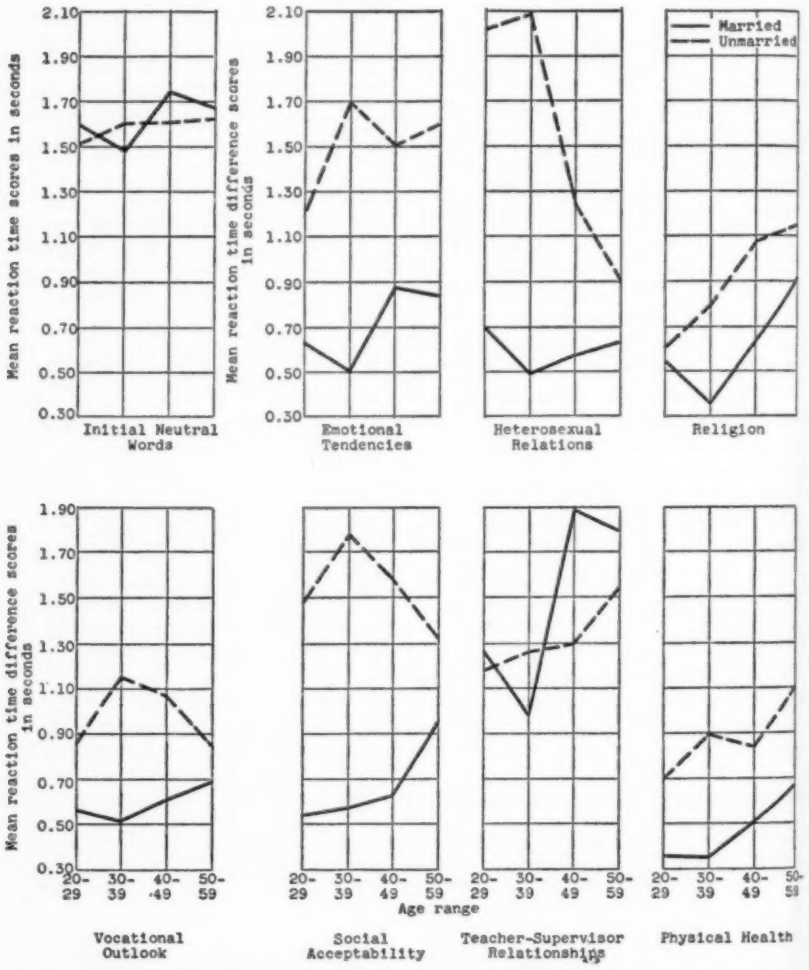

FIG. 1. Mean reaction time scores in seconds of married and unmarried teachers at various age ranges to initial neutral words, and mean reaction time difference scores in seconds of married and unmarried teachers at various age ranges to "critical" words taken from various psychological adjustment areas.

the "critical" words pertaining to the various psychological adjustment areas for the various age and marital groups are presented in graphic form in Fig. 1 in order to show the contrast visually. The means, standard deviations, and the results of the *d* tests are presented in Tables 3, 4, and 5.

Figure 1 shows a striking contrast between the married and unmarried women in their responses to words relating to Emotional Tendencies, with the unmarried responding so as to suggest greater emotional disturbance. As the *d*'s in Table 4 indicate, the differences were significant at all ages except 40 to 49. Although the general trend for both groups with increased age was in the direction of greater emotional disturbance, when the adjacent age groups were compared only the difference between the 30–39 and 40–49 groups in the case of the married women was significant. The greater general emotional disturbance on the part of the single woman may be explained either in terms of selection (being less likely to marry) or in terms of general frustration arising from their continued single state. Other studies suggest that marriage and children constitute one of the major sources of life satisfaction among women, and it is possible that frustration in this area might increase general emotionality especially in the 30's when a reorientation of goals (from marriage to career) is taking place.

That heterosexuality does contribute a

TABLE 3

MEANS AND STANDARD DEVIATIONS OF THE DIFFERENCE SCORES OF THE VARIOUS AREAS OF PSYCHOLOGICAL ADJUSTMENT

	20-29		30-39		40-49		50-59	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Emotional Tendencies								
Married	.625	.395	.509	.245	.882	.629	.838	.583
Unmarried	1.196	.619	1.711	1.343	1.512	1.463	1.602	1.176
Heterosexual Relations								
Married	.695	.330	.487	.297	.574	.331	.626	.432
Unmarried	2.022	1.081	2.075	2.065	1.226	.901	.877	.433
Religion								
Married	.550	.434	.360	.318	.655	.503	.917	.800
Unmarried	.614	.367	.806	.348	1.080	1.098	1.150	.940
Vocational Outlook								
Married	.563	.498	.515	.461	.609	.388	.696	.486
Unmarried	.857	.625	1.163	.594	1.076	.998	.858	.590
Social Acceptability								
Married	.543	.380	.575	.272	.628	.343	.966	1.002
Unmarried	1.486	1.176	1.780	1.185	1.578	.788	1.318	.948
Teacher-Supervisor Relations								
Married	1.264	.734	.978	.476	1.889	1.873	1.795	1.223
Unmarried	1.177	.698	1.264	.651	1.289	.890	1.540	.781
Physical Health								
Married	.364	.375	.349	.315	.514	.381	.676	.527
Unmarried	.689	.594	.899	.871	.848	.627	1.116	.753

TABLE 4

d TEST OF STATISTICAL SIGNIFICANCE OF DIFFERENCES IN THE DIFFERENCE SCORES BETWEEN MARRIED AND UNMARRIED FEMALES AT VARIOUS AGE LEVEL TO WORDS DRAWN FROM THE VARIOUS PSYCHOLOGICAL ADJUSTMENT AREAS (VALUES OF FISHER-BEHRENS *d*)

Adjustment Area	Age Ranges in Years			
	20-29	30-39	40-49	50-59
Emotional Tendencies	3.890**	4.405**	1.979	2.912*
Heterosexual Relations	5.869**	3.805**	3.396**	2.051
Religion	.563	4.728**	1.760	.943
Vocational Outlook	1.839	4.309**	2.180*	1.060
Social Acceptability	3.815**	4.955**	5.526**	1.276
Teacher-Supervisor Relationships	.429	1.774	1.447	.879
Physical Health	2.315*	2.971**	2.575*	2.394*

* .05 level of confidence.
** .01 level of confidence.

TABLE 5

d TEST OF STATISTICAL SIGNIFICANCE OF AGE RANGE DIFFERENCES IN THE DIFFERENCE SCORES IN THE VARIOUS AREAS OF PSYCHOLOGICAL ADJUSTMENT, BY MARITAL STATUS (VALUES OF FISHER-BEHRENS *d*)

Stimulus Area	Marital Status	20-29 vs. 30-39	30-39 vs. 40-49	40-49 vs. 50-59
Emotional Tendencies	Married	1.249	2.763**	.257
	Unmarried	1.742	.501	.240
Heterosexual Relations	Married	2.343*	.978	.478
	Unmarried	.114	1.884	1.745
Religion	Married	1.764	2.477*	1.386
	Unmarried	1.899	1.190	2.421*
Vocational Outlook	Married	.354	.780	.700
	Unmarried	1.774	.374	.940
Social Acceptability	Married	.342	.605	1.596
	Unmarried	.881	.710	1.054
Teacher-Supervisor Relationships	Married	1.635	2.357**	.210
	Unmarried	.456	.113	1.060
Physical Health	Married	.153	1.669	1.244
	Unmarried	.996	.238	1.368

* .05 level of confidence.
** .01 level of confidence.

major source of tension for single women, especially in the younger adult years, is clearly evident in the chart. At all ages except the oldest, the difference between the married and unmarried are highly significant. In comparisons of adjacent age groups, however, only the 30–39 vs. 40-49 difference in the case of single women is reliable, though the trend is obvious. These findings may be interpreted as sug-

gesting that the opportunity for heterosexual relations in marriage reduces this as a tension area for married women, while an eventual turning to a career may serve to remove heterosexuality from a central position and thus reduce it as a source of tension.

Religion, contrary to our hypothesis, also contributes a source of tension for both groups with the greatest difference between the marital status groups occurring at age 30–39 with the unmarried women showing significantly greater tension. At the adjacent ages of 30–39 vs. 40–49 married women showed significantly greater disturbance at the 40–49 level, while for single women the greatest disturbance appeared at the 40–49 vs. 50–59 year levels with the 50–59 level being the highest. It is difficult to discuss these findings, since the reasons for religious disturbance are not so readily apparent as in the case of the other areas in this study. Possibly the greater disturbance of 40–49-year-old married women is associated with raising children and orienting them to religion. For the single 30–39-year-old, it may represent an attempt to find a substitute for marriage or heterosexuality, while the 50–59 yr. disturbance may indicate a last major attempt to find in religion a source of satisfaction.

The area of Vocational Outlook results indicate that single women are significantly more disturbed about vocations during the 30–39 and 40–49 yr. levels than the married women are. Adjacent age group data demonstrate no significant differences in disturbance. Since the job is usually secondary to family life for the married women, her lack of concern in this area is easily understood. To the single woman the job is all important, particularly during the 30's and 40's when marriage becomes increasingly remote. Further, it is during the 30's that the job becomes a necessity, even if it is a job which is not satisfying to the individual, but one in which she must remain for economic reasons.

Figure 1 shows striking differences between married and unmarried women in the responses to words relating to Social Acceptability with the differences being very significant at all ages except the last. The adjacent age data show no significant differences between age levels for either group. Since marriage is in itself a form of social acceptance, this finding is not surprising. Apparently, the single female continues to strive for social acceptance well into her 40's, but by 50–59 either has found friends or has given up her search.

The area of Teacher-Supervisor Relationships offered results somewhat contrary to those hypothesized. There was no significant differences between the married and unmarried women at the age ranges studied. The only significant difference in the adjacent ranges was between 30–39 vs. 40–49 for married women with the disturbance greatest at the 40–49 yr. level. It had been expected that single women would have more conflict since they needed the job more. Apparently, however, neither group has any serious difficulty in this area except for the age mentioned above. At 40–49, the married women may feel more independent about their jobs, and may be more willing to resist authority, or may in fact be striving to reach a position of more authority.

The data indicate that single women are significantly more concerned about Physical Health at all ages studied, but there were no significant adjacent age differences for either group. This greater concern demonstrated by single women is probably due to their need to care for themselves. Where the married woman has others around to care for her, the single woman is alone. Poor physical health for the single woman may also relate to fear that illness may bring about loss of pay, which she can poorly afford.

Results of the Qualitative Word Analyses

The results of the qualitative analysis of the verbal responses to "critical" stimu-

TABLE 6

PERCENTAGES OF "DISTURBED" RESPONSES OF MARRIED (M) AND UNMARRIED (U) TEACHERS AT VARIOUS RANGES TO "CRITICAL" WORDS IN THE VARIOUS AREAS OF PSYCHOLOGICAL ADJUSTMENT

Adjustment Area	Age Range							
	20-29		30-39		40-49		50-59	
	M	U	M	U	M	U	M	U
Emotional Tendencies	7%	12%	8%	13%	12%	15%	14%	16%
Heterosexual Relations	2	7	6	15	10	16	3	9
Religion	6	11	3	9	2	7	8	4
Vocational Outlook	11	8	6	5	4	10	4	7
Social Acceptability	10	15	15	18	12	18	12	19
Teacher-Supervisor Relationships	6	6	6	9	10	6	5	6
Physical Health	4	6	3	6	6	7	3	11

lus words related to the various areas of psychological adjustment do not substantiate the hypotheses as do the results of the reaction time study technique described earlier. However, certain trends may be apparent from an inspection of Table 6 in which the percentages of "critical" words related to the various areas of psychological adjustment are presented. Since no significant differences were found, no conclusion can be reached.

SUMMARY

Reaction times of 25 single and 25 married elementary school teachers in each of four age groups (20–29, 30–39, 40–49, 50–59) were obtained to stimulus words representing seven areas of psychological adjustment. Differences between the married and single were especially large in the areas of Emotional Tendencies, Heterosexual Relations, Social Acceptability, and Physical Health, with the single women's responses reflecting greater tension in each of these areas. In two other areas marital status differences were reliable for only one or two age groups: Religion (30–39), Vocational Outlook (30–39, 40–49). In the remaining area (Teacher-Supervisor Relationships) no difference appeared at any age level. Except in the last instance, wherever differences appeared the single group's responses reflected more tension. Age trends were not clear-cut, but a trend toward greater tension with increasing age was apparent for both marital groups in Emotional Tendencies, Religion, Teacher-Supervisor Relationships, and Physical Health. A sharp drop among single women from the 30-39 age level onward was apparent in the case of Heterosexual Relations as a tension area.

These findings serve to identify sources of tension characterizing teacher personnel (and presumably other employed women) and thus are of interest to teachers themselves as well as to others concerned with the mental hygiene of teachers. Of special concern to administrators should be the finding that overall (ages and marital groups combined) the area of Teacher-Supervisor Relationships appeared to be a major source of tension among the several adjustment areas studied.

REFERENCES

BARKER, M. ELIZABETH. Summary of the reflection of personality adjustments of teachers to their efficiency in teaching. *J. educ. Res.*, **1948, 41,** 664–75.

FERRARO, C. D. Age and marital differences in degree of conflict within certain areas of psychological adjustment. Unpublished doctoral dissertation, Western Reserve Univer., 1957.

FISHER, R. A. The comparison of samples with possibly unequal variances. *Ann. Eugen.*, 1939, **9**, 174–80.

FISHER, R. A., & YATES, F. *Statistical tables for biological, agricultural and medical research*. New York: Hafner, 1953.

HARSH, C. M., & SCHRICKEL, H. G. *Personality: Development and assessment*. New York: Ronald, 1950.

HAVIGHURST, R. J. *Human development and education*. New York: Longmans, 1953.

HURLOCK, E. B. *Developmental psychology*. New York: McGraw-Hill, 1953.

JOHNSON, P. O. *Statistical methods in research*. New York: Prentice-Hall, 1949.

PECK, L. A study of the adjustment difficulties of a group of women teachers. *J. Educ. Psychol.*, 1936, **27**, 401–16.

POWELL, M. Age differences in degree of conflict within certain areas of psychological adjustment. Unpublished doctoral dissertation, Syracuse Univer., 1952.

POWELL, M. Age and sex differences in degree of conflict within certain areas of psychological adjustment. *Psychol. Monogr.*, 1955, **69**, No. 2.

RINSLAND, H. D. *A basic vocabulary of elementary school children*. New York: MacMillan, 1945.

SHARP, D. L. Group and individual profiles in the association-motor test. *Univer. Ia. Stud. Child Welf.*, 1937, **15**, 97–196.

SUKHATME, P. V. On Fisher and Behrens' test of significance for the differences in means of two normal samples. *Sankhya: Indian J. Stat.*, 1938, **4**, 39–48.

SYMONDS, P. M. The needs of teachers as shown in autobiographies. *J. Educ. Res.*, 1943, **36**, 662–77.

YOUNG, K. *Personality and problems of adjustment*. New York: Crofts, 1947.

(Received April 21, 1959)

Psychology and Mental Health

A report of the Institute on Education and Training for Psychological Contributions to Mental Health, held at Stanford University in August, 1955.

Edited by CHARLES R. STROTHER

Topics discussed at the Institute:

Training Needs of Psychologists in Community Mental Health

Specialization in Training

Practicum Training

Training for Therapy

Training for Research in the Mental Health Field

Problems of Departmental Organization

Price: $1.75

Order from:

American Psychological Association, Inc.
1333 Sixteenth Street, N. W.
Washington 6, D. C.

LOOKING FOR A BOOK PUBLISHER?

Book publication is one of the foundation stones of a professional career. To gain recognition, prestige, and advancement, it is often the means to an end.

Exposition Press offers a complete publishing service under our special academic imprint, *Exposition-University Books*. Many of our recent titles in the academic fields have gained popular reviews, professional recognition and steady sales. These books are the result of a publishing plan that is bringing a continuous flow of eminent writers to Exposition Press. They reflect the high editorial standards and quality of design and production which have won the respect of libraries, schools, booksellers and critics.

FREE! The behind-the-scenes story of book publishing revealed in two fact-filled, illustrated brochures, containing a detailed description of our 40% royalty-subsidy plan including a breakdown of contract terms and typical costs (in print for the first time). Copies are available on request. Your inquiries and manuscripts are invited. An editorial appraisal and sales evaluation will be furnished promptly without obligation.

Write Ely Pennington, Editorial Department,

EXPOSITION PRESS, *Main Office, 386 Park Ave. S., New York 16, N. Y.*

PSYCHOLOGY
——for Effective Teaching

George J. Mouly; University of Miami—While thoroughly covering traditional topics such as physical growth, emotional, social, and intellectual development, Professor Mouly adds an important dimension to his discussion by constantly applying psychological theory and knowledge of research results to specific teaching situations. Central to the author's point of view is the psychological principle that behavior is purposive. In Chapter 2, he presents clearly and simply the basic clinical concepts of needs, the self-image, and the phenomenal field. These concepts, which underlie the whole discussion, are expanded in subsequent chapters. *March 1960, 544 pp., $6.00*

The Adolescent: A Book of Readings, Revised
Jerome M. Seidman, Montclair State College
March 1960, 832 pp., $6.50

Inner Conflict and Defense
Daniel R. Miller and Guy E. Swanson, The University of Michigan
January 1960, 480 pp., $6.95

——————HENRY HOLT and Co., Inc., New York